UNIVERSITY OF CRETE

SCHOOL OF MEDICINE

PHD THESIS

ARTIFICIAL INTELLIGENCE IN MEDICINE:

A CLINICAL DECISION-SUPPORT FRAMEWORK BASED ON MACHINE LEARNING, STATISTICAL MIXED-EFFECT MODELING AND DEFEASIBLE REASONING FOR RHEUMATOID ARTHRITIS LONG-TERM PROGNOSIS UNDER BIOLOGIC THERAPY

GENITSARIDI IRINI

Heraklion, Crete, Greece *December 2020*

Cite as: *Irini Genitsaridi, "Artificial Intelligence in Medicine: A Clinical Decision Support Framework based on Machine Learning, Statistical Mixed-Effect Modeling and Defeasible Reasoning for Rheumatoid Arthritis Long-term Prognosis Under Biologic Therapy" (Doctoral Dissertation), Medicine Department, University of Crete, Greece, 2020.*

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΙΑΤΡΙΚΗ ΣΧΟΛΗ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΣΤΗΝ ΙΑΤΡΙΚΗ:

ΕΝΑ ΠΛΑΙΣΙΟ ΛΟΓΙΣΜΙΚΟΥ ΥΠΟΣΤΗΡΙΞΗΣ ΚΛΙΝΙΚΩΝ ΑΠΟΦΑΣΕΩΝ ΒΑΣΙΣΜΕΝΟ ΣΕ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ, ΣΤΑΤΙΣΤΙΚΑ ΜΟΝΤΕΛΑ ΜΙΚΤΩΝ ΕΠΙΔΡΑΣΕΩΝ ΚΑΙ ΑΝΑΙΡΕΣΙΜΗ ΣΥΛΛΟΓΙΣΤΙΚΗ ΓΙΑ ΤΗ ΜΑΚΡΟΠΡΟΘΕΣΜΗ ΠΡΟΓΝΩΣΗ ΤΗΣ ΡΕΥΜΑΤΟΕΙΔΟΥΣ ΑΡΘΡΙΤΙΔΑΣ ΥΠΟ ΒΙΟΛΟΓΙΚΗ ΘΕΡΑΠΕΙΑ

ΓΕΝΙΤΣΑΡΙΔΗ ΕΙΡΗΝΗ

Ηράκλειο, Κρήτη, Ελλάδα *Νοέμβριος 2020*

Αναφορά ως: *Ειρήνη Γενιτσαρίδη, "Τεχνητή Νοημοσύνη στην Ιατρική: Ένα Πλαίσιο Λογισμικού Υποστήριξης Κλινικών Αποφάσεων βασισμένο σε Μηχανική Μάθηση, Στατιστικά Μοντέλα Μικτών Επιδράσεων και Αναιρέσιμη Συλλογιστική για τη Μακροπρόθεσμη Πρόγνωση της Ρευματοειδούς Αρθρίτιδας Υπό Βιολογική Θεραπεία" (Διδακτορική Διατριβή), Τμήμα Ιατρικής, Πανεπιστήμιο Κρήτης, Ελλάδα, 2020.*

Author: ……..……...… **Genitsaridi Irini**
Lead Data and Computer Scientist, ISeeChange Company, New Orleans, Louisiana, United States.

Three-Member Committee:

Main Supervisor: ……..……...….. **Sidiropoulos Prodromos**
Director of Rheumatology Clinic, University Hospital of Heraklion (PAGNI), Greece.
Professor of Rheumatology, Medical School, University of Crete (UOC), Greece.

Co-supervisor: ……...…….. **Plexousakis Dimitris**
Director of Computer Science Institute, Foundation of Research and Technology, Hellas, Greece.
Professor of Computer Science Department, University of Crete (UOC), Greece.

Co-supervisor: ……..……..... **Marias Konstantinos**
Associate Professor, Hellenic Mediterranean University, Crete (HMU), Greece.
Computational Biomedicine Laboratory, Foundation of Research and Technology, Hellas, Greece.

Seven-Member Committee:

Member 1 *(Three-Member Committee*): …...…….....…….…… **Sidiropoulos Prodromos**
Member 2 *(Three-Member Committee*): ……...…….. **Plexousakis Dimitris**
Member 3 *(Three-Member Committee*): ……....…… **Marias Konstantinos**

Member 4: ……..…….....…. **Bertsias George**
Adjunct Professor, Rheumatology-Clinical Immunology, Medical School University of Crete, Greece. Collaborating Researcher, Rheumatology Clinic, University Hospital of Heraklion (PAGNI), Greece.

Member 5: ……..…….....… **Koutroubakis Ioannis**
Associate Professor of Gastroenterology, Medical School, University of Crete, Greece.

Member 6: ……..…… **Potamias George**
Principal Researcher, Computer Science, Foundation of Research and Technology, Hellas, Greece.

Member 7: …… **Exarchos Themistoklis**
Adjunct Professor, Information Science and Informatics, Ioanian University, Corfu, Greece.

Συγγραφέας: ……..…….……...… **Γενιτσαρίδη Ειρήνη**

Επικεφαλής Επιστήμης Δεδομένων και Υπολογιστών, Εταιρεία ISeeChange, Νέα Ορλεάνη, Λουιζιάνα, Ηνωμένες Πολιτείες Αμερικής.

Τριμελής Εισηγητική Επιτροπή:

Κύριος Επιβλέπων: ……..……..…….…….... **Πρόδρομος Σιδηρόπουλος**

Διευθυντής Ρευματολογικής Κλινικής, Πανεπιστημιακό Γενικό Νοσοκομείο Ηρακλείου (ΠΑΓΝΗ), Κρήτη. Καθηγητής Ρευματολογίας, Ιατρική Σχολή, Πανεπιστήμιο Κρήτης.

Συν-επιβλέπων: ……..……..…….. **Πλεξουσάκης Δημήτρης**

Διευθυντής του Ινστιτούτου Πληροφορικής, Ίδρυμα Τεχνολογίας και Έρευνας (ΙΤΕ), Κρήτη.
Καθηγητής, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης.

Συν-επιβλέπων: ……..……..……..……..... **Μαριάς Κωνσταντίνος**

Αναπληρωτής Καθηγητής, Ελληνικό Μεσογειακό Πανεπιστήμιο, Κρήτη. Εργαστήριο Υπολογιστικής Βιοϊατρικής, Ινστιτούτο Πληροφορικής, Ίδρυμα Τεχνολογίας και Έρευνας, Κρήτη.

Εφταμελής Εισηγητική Επιτροπή:

Μέλος 1 (*Τριμελής Επιτροπή*): …...…….....…….…. **Πρόδρομος Σιδηρόπουλος**
Μέλος 2 (*Τριμελής Επιτροπή*): ……..……..…….. **Πλεξουσάκης Δημήτρης**
Μέλος 3 (*Τριμελής Επιτροπή*): ……...……...… **Μαριάς Κωνσταντίνος**

Μέλος 4: ……...……...……...… **Μπερτσίας Γεώργιος**

Επίκουρος Καθηγητής Ρευματολογίας-Κλινικής Ανοσολογίας, Ιατρική Σχολή, Πανεπιστήμιο Κρήτης. Συνεργαζόμενος Ερευνητής, Ρευματολογική Κλινική, Πανεπιστημιακό Γενικό Νοσοκομείο Ηρακλείου (ΠΑΓΝΗ), Κρήτη.

Μέλος 5: ……..…… **Κουτρουμπάκης Ιωάννης**

Αναπληρωτής Καθηγητής Γαστρεντερολογίας, Ιατρική Σχολή, Πανεπιστήμιο Κρήτης.

Μέλος 6: ……..…….....…**Ποταμιάς Γεώργιος**

Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, Ίδρυμα Τεχνολογίας και Έρευνας (ΙΤΕ), Κρήτη.

Μέλος 7: …….....**Έξαρχος Θεμιστοκλής**

Επίκουρος Καθηγητής, Τμήμα Επιστήμης Πληροφορίας και Πληροφορικής, Ιόνιο Πανεπιστήμιο, Κέρκυρα.

Acknowledgments

I would like to thank my supervisor, Mr. Prodromos Sidiropoulos (Professor of Rheumatology, Medical School, University of Crete and Director of Rheumatology Clinic, General University Hospital of Heraklion, Crete - PAGNI), for his guidance, his valuable advice, the effort and time he devoted in the design and implementation of this study. This work would not be possible without the support and trust he has shown in our collaboration, despite the inevitable difficulties of its interdisciplinary nature. I would like to thank him for the opportunity to become a member of a productive interdisciplinary team!

I would also like to thank the collaborating researcher of the rheumatology clinic of PAGNI, Irini Flouri, for her valuable advice, positive attitude and significant assistance in all stages of this study, from design to implementation. I also thank Mr. Plexousakis Dimitris and Mr. Marias Kostas for their comments and advice as members of the advisory committee and the internal and external collaborators of the Rheumatology Clinic of PAGNI, Bertsias George, Boumbas Dimitris, Drosos Alexandros, Boki Kyriaki and Skopouli Fotini for their comments upon publication of the results.

Many thanks to the patients who contributed with their data to this study, as well as the medical and nursing staff of the collaborating centers for their contribution to the collection and management of these data. The research work was funded in 1/7/2017 – 31/12/2018 by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI PhD Fellowship grant (no 1722) 2018.

On a personal level, warm thanks to my friends for the psychological support, the positive mood and the optimism they have provided to me.

Finally, I thank and dedicate this study to my wonderful parents, Kostas Genitsaridis and Stella Kokkinou, to my amazing sisters, Eleni and Gianna (sternoulia) and to my beloved life companion, Michalis Orphanoudakis, for their continuous support in every step of my life.

Thank you!

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου, κύριο Πρόδρομο Σιδηρόπουλο (Καθηγητή της Ιατρικής Σχολής Πανεπιστήμιου Κρήτης και Διευθυντή της Ρευματολογικής Κλινικής του Δημόσιου Πανεπιστημιακού Γενικού Νοσοκομείου Κρήτης ΠΑΓΝΗ) για την καθοδήγηση, τις πολύτιμες συμβουλές του, την προσπάθεια και το χρόνο που αφιέρωσε στην σχεδίαση και υλοποίηση αυτής της μελέτης. Αυτή η μελέτη δεν θα ήταν δυνατόν να πραγματοποιηθεί χωρίς τη στήριξη και εμπιστοσύνη που έδειξε στη συνεργασία μας, παρά τις αναπόφευκτες δυσκολίες της διεπιστημονικής της φύσης. Θα ήθελα να τον ευχαριστήσω για την ευκαιρία να αποτελέσω μέλος μιας παραγωγικής διεπιστημονικής ομάδας!

Θα ήθελα να ευχαριστήσω ακόμα, την συνεργαζόμενη ερευνήτρια της ρευματολογικής κλινικής του ΠΑΓΝΗ, Ειρήνη Φλουρή, για τις συμβουλές της, την θετική της διάθεση και την σημαντική βοήθεια που παρείχε σε όλα τα στάδια αυτής της μελέτης, από την σχεδίαση μέχρι την υλοποίηση. Ευχαριστώ επίσης τον κύριο Πλεξουσάκη Δημήτρη και στον κύριο Μαριά Κώστα για τις παρατηρήσεις και συμβουλές τους ως μέλη της εισηγητικής επιτροπής και τους εσωτερικούς και εξωτερικούς συνεργάτες της Ρευματολογικής κλινικής του ΠΑΓΝΗ, Μπερτσιά Γιώργο, Μπούμπα Δημήτρη, Δρόσο Αλέξανδρο, Μποκή Κυριακή, Σκοπούλη Φωτεινή για τα εύστοχα σχόλια τους κατά την δημοσίευση των αποτελεσμάτων.

Πολλά ευχαριστώ στους ασθενείς που συνέβαλλαν με τα δεδομένα τους σε αυτή τη μελέτη αλλά και το ιατρικό και νοσηλευτικό προσωπικό των συνεργαζόμενων κέντρων για την συμβολή τους στην συλλογή και διαχείριση αυτών των δεδομένων. Η ερευνητική εργασία χρηματοδοτήθηκε τη χρονική περιόδο 1/7/2017 – 31/12/2018 από το Ελληνικό Ίδρυμα Έρευνας και Καινοτομίας (ΕΛ.ΙΔ.Ε.Κ.) και τη Γενική Γραμματεία Έρευνας και Τεχνολογίας (ΓΓΕΤ), στο πλαίσιο της επιχορήγησης διδακτορικών υποτροφιών ΕΛ.ΙΔ.Ε.Κ (αρ. 1722).

Σε προσωπικό επίπεδο, θερμές ευχαριστίες στους φίλους μου για την ψυχολογική υποστήριξη, την θετική διάθεση και αισιοδοξία που μου μετέφεραν όλον αυτό το καιρό.

Τέλος, ευχαριστώ και αφιερώνω αυτή τη μελέτη στους υπέροχους γονείς μου Κώστα Γενιτσαρίδη και Στέλλα Κόκκινου, στις καταπληκτικές αδερφές μου Ελένη και Γιάννα και στον αγαπημένο μου

σύντροφο Μιχάλη Ορφανουδάκη για την στήριξη και συμπαράσταση που μου παρείχαν και εξακολουθούν να μου παρέχουν σε κάθε βήμα της ζωής μου.

Σας ευχαριστώ!

To my family!

Στην οικογένεια μου!

Περίληψη

Η *τεχνητή νοημοσύνη* (AI), είναι ο κλάδος πληροφορικής που ασχολείται με τη σχεδίαση και την υλοποίηση ευφυών συστημάτων ικανά να συλλέγουν πληροφορίες, να εκτελούν ανάλυση δεδομένων και να εφαρμόζουν κατάλληλες ενέργειες, για να καλύψουν τις ανάγκες διαφόρων περιβαλλόντων, μιμούμενα την ανθρώπινη συμπεριφορά ως προς την συλλογιστική, μάθηση και επίλυση προβλημάτων. Η εφαρμογή της τεχνητής νοημοσύνης στην ιατρική επιτρέπει τη συλλογή και ανάλυση ιατρικών πληροφοριών και την εφαρμογή κατάλληλων δράσεων, για την υποστήριξη της πρόληψης, της διάγνωσης, της θεραπείας και της πρόγνωσης των ασθενειών. Τα κλινικά συστήματα υποστήριξης αποφάσεων (CDS) είναι συστήματα που υποστηρίζουν τη διαδικασία λήψης αποφάσεων των ιατρών. Τα συστήματα CDS είναι σε θέση να πραγματοποιήσουν σύνθετες αναλύσεις ιατρικών δεδομένων, να αναγνωρίσουν συσχετισμούς στα ιατρικά δεδομένα, να προσομοιώσουν και να βελτιώσουν τη διαδικασία ιατρικής συλλογιστικής και να παρέχουν προγνωστικές πληροφορίες που βοηθούν στον έγκαιρο εντοπισμό κρίσιμων καταστάσεων. Η παρούσα διατριβή επικεντρώνεται στην ανάπτυξη και παρουσίαση του κλινικού συστήματος υποστήριξης αποφάσεων CDS-RA για την διαχείριση και μακροπρόθεσμη πρόγνωση ασθενών με ρευματοειδή αρθρίτιδα (RA) υπό βιολογική θεραπεία. Το σύστημα, CDS-RA χρησιμοποιεί μεθόδους τεχνητής νοημοσύνης για τη διεξαγωγή προηγμένων αναλύσεων ιατρικών δεδομένων βασισμένων σε στατιστικά μοντέλα μεικτών επιδράσεων, σε μοντέλα μηχανικής μάθησης και σε πολιτικές εξαγωγής συμπερασμάτων από λογικούς ιεραρχημένους κανόνες βασισμένους στην αναιρέσιμη συλλογιστική.

Ένας σημαντικός στόχος του CDS-RA συστήματος είναι να παρέχει προγνωστική λειτουργικότητα ικανή για την πρώιμη πρόβλεψη και αιτιολόγηση του επιπέδου επίμονης νόσου ενός ασθενούς με ρευματοειδή αρθρίτιδα υπό βιολογική θεραπεία. Το επίπεδο επίμονης νόσου (PDL) ορίστηκε ως το ίδιο επίπεδο δραστηριότητας της νόσου (DAS28 εντός συγκεκριμένου εύρους), που παρουσιάστηκε σε έναν ασθενή υπό βιολογική θεραπεία, για τουλάχιστον το ήμισυ της 5ετούς κλινικής παρακολούθησης, σωρευτικά και ανεξάρτητα από τις διακυμάνσεις. Ορίστηκαν τρία επίπεδα "επίμονης" νόσου (PDL), το επίπεδο LDA (DAS28≤3.2), MDA (3.2 <DAS28≤5.1) και HDA (DAS28> 5.1), αντίστοιχα. Η διατριβή παρέχει στοιχεία για την κλινική σημασία της πρώιμης κατηγοριοποίησης των ασθενών στα επίπεδα PDL αναλύοντας τη σχέση τους με διαφορετικά μακροπρόθεσμα αποτελέσματα. Για τις αναλύσεις ανακτήθηκαν δεδομένα ασθενών

από το ελληνικό πολυκεντρικό Ελληνικό Μητρώο Βιολογικών Θεραπειών (HeRBT) από 8 νοσοκομεία της Ελλάδας. Πραγματοποιήθηκε συγκριτική ανάλυση μεταξύ ασθενών διαφορετικών επιπέδων επίμονης νόσου (PDL) στα ακόλουθα μακροπρόθεσμα αποτελέσματα, (α) στην πενταετή πορεία λειτουργικότητας (HAQ) των ασθενών και (β) στο σύνολο σοβαρών ανεπιθύμητων συμβάντων στα 5 χρόνια βιολογικής θεραπείας. Αναπτύχθηκε ένα πολυπαραγοντικό μοντέλο μεικτών επιδράσεων (mixed-effect model) με βάση τις πενταετές πορείες λειτουργικότητας των ασθενών που έδειξε ότι οι ασθενείς στο επίπεδο MDA συσχετίστηκαν με χειρότερη πορεία λειτουργικότητας 5 ετών από τους ασθενείς στο επίπεδο LDA (+0.27 υψηλότερη τροχιά HAQ στο MDA από το LDA, $p<0.0001$) ενώ οι ασθενείς του επιπέδου HDA συσχετίστηκαν με ακόμη υψηλότερο περιορισμό λειτουργικότητας 5 ετών από τους ασθενείς επιπέδου LDA (+0.69 υψηλότερη τροχιά HAQ στο HDA από το LDA, $p<0.0001$). Παρόμοια διαφοροποίηση παρουσίασαν τα σοβαρά ανεπιθύμητα συμβάντα (0.2±0,48 σε LDA, 0.5±0.96 σε MDA και 0.89±1.7 σε HDA, $p<0.01$). Το CDS-RA απεικονίζει τις διαφοροποιημένες πορείες 5-ετής λειτουργικότητας και σοβαρών ανεπιθύμητων συμβάντων των ομάδων ασθενών σε LDA, MDA και HDA.

Η ομάδα ασθενών με ρευματοειδή αρθρίτιδα στο επίπεδο MDA (επίμονη μέτρια νόσος) που δεν βελτιώνει ούτε επιδεινώνεται (εκτός της μέτριας ενεργότητας νόσου) υπό βιολογική θεραπεία για σημαντικό χρονικό διάστημα, δεν έχει ερευνηθεί επαρκώς στη βιβλιογραφία. Έτσι, η διατριβή επικεντρώθηκε επίσης στην ανάλυση αυτής της ομάδας και συγκεκριμένα στην εσωτερική της ετερογένεια. Οι ασθενείς με MDA υπο-κατηγοριοποιήθηκαν σε δύο υπο-ομάδες χαμηλότερης και υψηλότερης MDA. Ένα πολυπαραγοντικό μοντέλο μεικτών επιδράσεων (mixed-effect model) αναπτύχθηκε με βάση τις πενταετές πορείες λειτουργικότητας των ασθενών που έδειξε ότι η υπό-ομάδα ασθενών σε υψηλότερη MDA συσχετίστηκε με χειρότερη πορεία λειτουργικότητας 5 ετών από την υπό-ομάδα ασθενών σε χαμηλότερη MDA (+0.26 υψηλότερη τροχιά HAQ στην υψηλότερη MDA, $p <0.0001$). Παρόμοια διαφοροποίηση παρουσίασαν τα σοβαρά ανεπιθύμητα συμβάντα (0.32±0.6 στην χαμηλότερη MDA και 0.64±1.16 στην υψηλότερη MDA, $p=0.038$). Η ετερογένεια που βρέθηκε μεταξύ ασθενών με χαμηλότερη και υψηλότερη MDA μπορεί να βοηθήσει μελλοντικές στρατηγικές στοχευμένων θεραπειών (T2T) ώστε να προσαρμόσουν τις θεραπείες για αυτές τις υπο-ομάδες προκειμένου να βελτιώσουν τα μακροπρόθεσμα αποτελέσματά τους.

Το σύστημα CDS-RA που αναπτύχθηκε περιλαμβάνει λειτουργικότητα τεχνητής νοημοσύνης (AI) που υποστηρίζει την πρόγνωση του επιπέδου επίπονης νόσου (LDA, MDA και HDA) ενός ασθενή υπό βιολογική θεραπεία, βασισμένη σε τρεις στοιχειοθετημένες πολιτικές διαφορετικής προτεραιότητας. Η πρώτη πολιτική $1^{ης}$ προτεραιότητας κατηγοριοποιεί τον ασθενή σε

ένα επίπεδο επίπονης νόσου όταν ο ασθενής πληρεί τα κριτήρια κάποιου επιπέδου με βάση τον ορισμό των επιπέδων από υπάρχοντα μακροχρόνια δεδομένα της πορεία της νόσου του. Η δεύτερη πολιτική 2^{ης} προτεραιότητας αφορά ασθενείς για τους οποίους δεν υπάρχουν μακροχρονία δεδομένα της πορείας της νόσου τους ώστε να πληρούν τον ορισμό κάποιο επιπέδου αλλά υπάρχει εκτίμηση ειδικευμένου ιατρού που τους παρακολουθεί για το επίπεδο επίμονης νόσου τους. Η τρίτη πολιτική 3^{ης} προτεραιότητας αφορά ασθενείς για τους οποίους δεν υπάρχουν μακροχρόνια δεδομένα της πορείας της νόσου τους ώστε να πληρούν τον ορισμό κάποιο επιπέδου και ούτε υπάρχει εκτίμηση ειδικευμένου ιατρού για το επίπεδο επίμονης νόσου τους αλλά υπάρχουν πρώιμα δεδομένα από τους αρχικούς (6-9) μήνες της θεραπείας τους. Η πολιτική αυτή χρησιμοποιεί προβλεπτικά μοντέλα μηχανικής μάθησης που αναπτύχθηκαν στο σύστημα για την εξατομικευμένη πρόβλεψη του επιπέδου επίμονης νόσου ενός ασθενή από πρώιμα δεδομένα θεραπείας. Οι τρείς προγνωστικές πολιτικές και η ιεράρχηση τους εκφράστηκαν με λογικούς κανόνες αναιρέσιμης συλλογιστικής και εισάχθηκαν ως θεωρία σε ένα εργαλείο συλλογισμού κανόνων (rule-based reasoning engine) ενσωματωμένο στο σύστημα CDS-RA για να υποστηρίζεται η εξαγωγή συμπερασμάτων. Η AI θεωρία λογικών κανόνων είναι προσβάσιμη, επαναχρησιμοποιήσιμη, διαμορφώσιμη και επεκτάσιμη για την υποστήριξη πρόσθετων ιεραρχημένων ιατρικών πολιτικών.

Δύο πολυπαραγοντικά προβλεπτικά μοντέλα λογιστικής παλινδρόμησης (logistic regression) από το πεδίο της μηχανικής μάθησης αναπτύχθηκαν στο σύστημα για την εξατομικευμένη πρόβλεψη του επιπέδου επίμονης νόσου ενός ασθενή (LDA, MDA ή HDA) από πρώιμα δεδομένα θεραπείας. Το πρώτο ανέδειξε ως πρώιμους δείκτες LDA σε σχέση με τα υπόλοιπα επίπεδα, το ανδρικό φύλο (OR 0.38 για γυναικείο φύλο, p=0.02), την χαμηλότερη ενεργότητα ασθένειας (OR 0.42 για DAS28 ανά μονάδα, p=0.001) και λειτουργικότητα (OR 0.3 για HAQ ανά μονάδα, p=0.01) στην έναρξη της θεραπείας και την χαμηλότερη μέση ενεργότητα ασθένειας το πρώτο εξάμηνο (OR 0.2 για DAS28 ανά μονάδα, p<0.001). Το δεύτερο ανέδειξε ως πρώιμους δείκτες MDA σε σχέση με HDA στην έναρξη της θεραπείας, την μικρότερη ηλικία (OR 1.04 ανά 1 χρόνο, p=0.003), την μικρότερη διάρκεια νόσου από 2 χρόνια (OR 2.65, p=0.026), την χρήση πρεδνιζολόνης (OR 1.81, p=0.033), την χαμηλότερη ενεργότητα ασθένειας (OR 0.56 για DAS28 ανά μονάδα, p<0.001) και την λειτουργικότητα (OR 0.21 για HAQ ανά μονάδα, p<0.001), ενώ πρώιμοι δείκτες του πρώτου εξάμηνου θεραπείας ήταν η χαμηλότερη μέση ενεργότητα ασθένειας (OR 0.42 για DAS28 ανά μονάδα, p<0.001), η βελτιωμένη μέση λειτουργικότητα σε σχέση με την έναρξη (OR 2.89, p=0.002) και η απουσία σοβαρών ανεπιθύμητων συμβάντων (OR 0.32, p=0.047).

Συνολικά, το CDS-RA παρέχει ένα τεχνολογικό περιβάλλον τεχνητής νοημοσύνης με ευρύ φάσμα λειτουργικών υπηρεσιών, συμβατό για χρήση από κινητές συσκευές (tablet), που

υποστηρίζει τη διαχείριση της θεραπείας ασθενών με ρευματοειδή αρθρίτιδα, διευκολύνει την αλληλεπίδραση ασθενούς και κλινικού ιατρού και παρέχει εξατομικευμένες προγνωστικές πληροφορίες για την μακροπρόθεσμη έκβαση των ασθενών υπό βιολογική θεραπεία. Οι καινοτόμες υπηρεσίες του συστήματος χρησιμοποιούν πολυπαραγοντικά μοντέλα μεικτών επιδράσεων, προβλεπτικά μοντέλα μηχανικής μάθησης και αναιρέσιμη συλλογιστική, για να παρέχουν πολύτιμες προγνωστικές πληροφορίες κατά την διαδικασία κλινικών αποφάσεων. Το σύστημα CDS-RA στοχεύει στην υποστήριξη των κλινικών ιατρών κατα την βιολογική θεραπεία ασθενών με ρευματοειδή αρθρίτιδα ώστε να βελτιωθούν τα αποτελέσματα τους.

Table of Contents

List of Tables

List of Figures

Abstract

Artificial intelligence (AI), is state-of-art information technology that provides intelligent software frameworks, able to collect information, perform data analysis and implement appropriate actions, to meet the needs of various environments, mimicking human behavior in reasoning, learning and problem resolution. The application of AI in medicine enables the collection and analysis of medical information and the implementation of appropriate actions, to support disease prevention, diagnosis, therapy and prognosis. Clinical decision-support systems (CDS) are systems that support the clinicians' decision-making process. CDS systems are able to perform complex medical data analyses, unravel medical data associations, simulate and enhance the medical reasoning process and support prognostic information which assists the identification of critical situations. The present thesis focuses on the development of a clinical decision-support system (CDS-RA) to support Rheumatoid Arthritis (RA) management and long-term prognosis under biologic therapy. The CDS-RA utilizes artificial intelligence methods to conduct advanced medical data analyses based on statistical mixed-effect models, machine learning and defeasible reasoning.

An important objective of the CDS-RA is to provide prognostic functionality able to early predict and reason about the persistent disease level of a RA patient under biologic therapy. Persistent disease level (PDL) was defined for patients under biologic therapy as the same disease activity level (DAS28 within a specific range) for at least half of the 5-year clinical follow-up, cumulatively and irrespective of fluctuations. Three PDL patient groups were specified, the LDA (DAS28$\leq$3.2), MDA (3.2<DAS28$\leq$5.1) and HDA (DAS28>5.1) groups, respectively. The thesis provides evidence on the clinical importance of early patient categorization into the PDL groups by analyzing their association with different long-term outcomes. Patients' data required for the analyses were retrieved from the Greek nationwide multicenter registry HeRBT (Hellenic Registry of Biologic Therapies) of seven healthcare centers in Greece. Two patient outcomes were compared between the groups, (a) the 5-year functionality trajectories and (b) the serious adverse events (SAEs) at 5 years of biologic therapy. A multivariable mixed-effect model was developed based on patients' 5-year functionality trajectories which showed that MDA was associated with worse 5-year functionality course than LDA group (+0.27 higher HAQ trajectory in MDA than LDA, p<0.0001) and also HDA was associated with even higher 5-year functionality limitation than the LDA group (+0.69 higher HAQ trajectory in HDA than LDA, p<0.0001). Similarly, SAEs were differentiated (0.2±0.48 in LDA, 0.5±0.96 in MDA and 0.89±1.7 in HDA;

p<0.01). The CDS-RA system provides a functional service that depicts the differentiated 5-year group trajectories of functionality and serious adverse events.

The MDA patient group that under biologic therapy neither improves nor deteriorates (outside moderate disease activity) for a significant amount of time, is an under-researched group in RA literature. Thus, the thesis also focused on the analysis of this group and in particular its internal heterogeneity. Specifically, MDA patients were sub-categorized into two subgroups of lower and higher MDA. A multivariable mixed-effect model was developed based on patients' 5-year functionality trajectories which showed that the higher MDA subgroup was associated with worse 5-year functionality course than the lower MDA subgroup (+0.26 higher HAQ trajectory in higher-MDA, p<0.0001). Similarly, SAEs were differentiated (0.32 ±0.6 in lower MDA and 0.64 ±1.16 in higher MDA; p=0.038). The heterogeneity found between lower and higher MDA patients can assist future T2T strategies to tailor treatments for these subgroups in order to improve their outcomes.

The CDS-RA system includes an AI Layer that supports a prognostic functional service of patient PDL group (LDA, MDA, and HDA) based on three policies that utilize different medical evidence sources in decreasing priority. The first policy of highest priority is based on long-term disease data when they exist for a specific patient. Specifically, the policy categorizes the patient into a PDL group when the patient's long-term follow-up fulfills the criteria membership for a PDL group (LDA, MDA, and HDA) by definition. The second policy is based on clinician's expert opinion for group membership when it is provided. The third policy is a predictive service developed in the CDS-RA system for early prediction of the patient PDL group when neither long-term patient data exists, nor can clinicians provide information on the long-term disease level course that a patient will develop. The prognostic policies of the AI Layer and their prioritization were expressed with defeasible logical rules and were loaded in a AI engine integrated in the CDS-RA system that supports rule-based reasoning in Defeasible Logic. The AI logical rule theory is accessible, reusable, configurable and extendable to support additional prioritized medical policies.

The predictive service of the AI Layer utilizes early (first 6 to 9 months) patient data in order to provide a personalized prediction of the patient's long-term persistent disease level (LDA, MDA, or HDA). Two multivariable logistic regression Machine Learning models were developed. The first model yielded, males (OR 0.38 for females, p=0.02), lower baseline disease activity (OR 0.42 for DAS28 per unit, p=0.001), lower baseline functionality (OR 0.3 for HAQ per unit, p=0.01) and lower first semester's average disease activity (OR 0.2 for DAS28 per unit, p<0.001) as early predictors for LDA compared to other groups. The

second model yielded, younger age (OR 1.04 per year, p=0.003), shorter disease duration (OR 2.65 for duration years<2, p=0.026), prednisolone initiation at baseline (OR 1.81, p=0.033), lower baseline disease activity (OR 0.56 for DAS28 per unit, p<0.001), lower baseline functionality (OR 0.21 for HAQ per unit, p<0.001) and lower first semester's average disease activity (OR 0.42 for DAS28 per unit, p<0.001), first semester's average functionality improvement compared to baseline (OR 2.89, p=0.002) and lower occurrence of first semester's serious adverse events (OR 0.32 for SAEs count>0, p=0.047) as early predictors for MDA compared to HDA.

Overall, CDS-RA provides a state-of-art AI technological environment with a wide range of functional services that is mobile compatible, supports RA patient data management over time, facilitates patient-clinician interaction and provides personalized prognostic information for the long-term outcome of RA patients under biologic therapy. The innovative functionality integrates seamlessly statistical multivariable mixed-effect modeling, machine learning predictive modeling and defeasible logical reasoning to provide valuable insights during the clinical-decision making process. CDS-RA is aimed to assist clinicians in the biologic treatment of RA patients in order to support improved patient outcomes.

CHAPTER I. Artificial Intelligence in Medicine

Artificial intelligence (AI), is state-of-art information technology that provides intelligent software frameworks, able to collect information, perform data analysis and implement appropriate actions, to meet the needs of various environments [1]. The application of artificial intelligence in medicine (AIM) [2] enables the collection and analysis of medical information and the implementation of appropriate actions, to support disease prevention, diagnosis, therapy management and prognosis strategies.

Electronic health record systems (EHRs) have emerged as AIM technology that can support disease information recording and management, such as patient's medical history, symptoms, diagnosis, treatments and disease progression data [3]. EHRs provide intelligent services for both clinician and patient users to enhance their continuous communication and motivate patient active engagement in the therapy process. EHRs or their subparts that are targeted for patient use are called *personal health record systems* (PHRs). PHRs empower patients to make informed decisions and to provide frequent feedback on their health status [4, 5].

Sophisticated EHR solutions may support management and analysis of enormous medical data volumes (Big Data) [6] that require specialized technological approaches. In addition, they may include natural language processing (NLP) capabilities [7] in order to automatically identify clinically meaningful information (such as disease symptoms and comorbidities) from free-text descriptions of clinicians and patients.

AIM technological systems that support the clinician decision-making process are called *clinical decision-support systems* (CDS systems or CDSs) [8]. CDS systems are able to perform complex medical data analysis and are usually integrated in EHRs. Medical data associations and predictive information can be derived from the CDS systems and they are able to simulate and enhance the medical reasoning process [9, 10]. Patient prognosis and critical situations can also be identified from CDS systems and they enable appropriate prescriptive functionality such as intelligent alerts and recommendations.

The present thesis focuses on the development of a clinical decision-support system (CDS) to support Rheumatoid Arthritis (RA) long-term prognosis under biologic therapy. The CDS utilizes artificial intelligence methods for advanced data analysis such as statistical mixed-effect models, machine learning and defeasible reasoning. The rest of the thesis is structured as follows. Chapter I describes AI methods for advanced medical data analysis. Chapter II provides information on RA disease, related work on RA disease management and

prognosis as well as literature limitations on RA prognosis under biologic therapy. Chapter III introduces the CDS system and analyzes its functionality, development methodology and implementation results while Chapter IV discusses results, future directions and concludes the thesis.

1. Artificial Intelligence for Advanced Medical Data Analysis

Artificial intelligence technology supports advanced methods for medical data analysis to provide (a) descriptive, (b) predictive and (c) prescriptive information. *Descriptive information* includes data patterns, associations and data structures that describe the data. *Predictive information* includes predictions for the future that are derived from advanced algorithmic processes on the descriptive information and also include an estimation of the predictions uncertainty. *Prescriptive information* includes logical inferences that may represent meaningful conclusions, recommendations and actions to be taken, as the result of the utilization of the descriptive and predictive information, in simulated human reasoning processes.

Data analysis and by extension medical data analysis is performed to prepare data, to compare and explain them and to predict and reason about them. Methodologies of data analysis can be categorized into the following types, (a) *Preparatory Analysis* that prepares data and improves its quality with preprocessing, (b) *Association Analysis of Difference* that evaluates group data differences, (c) *Association Analysis of Correlation* that evaluates the correlation between data variables, (d) *Explanatory Association Analysis* that associates data with explanatory parameters, (e) *Predictive Analysis* that derives predictive information from data and (f) *Logical Analysis or Logical Reasoning* that provides conditional conclusions from data and simulates the human reasoning process. These types are complementary to each other since they address different aspects of data analysis. Hybrid methodologies can also be generated that integrate aspects from different analysis types.

Sophisticated AIM technological solutions are able to support many types of data analysis in complex computational processes (e.g. sequential, parallel, iterative and multilevel algorithmic flows) to meet specific medical domain needs. Figure 1 depicts an example of many data analyses in a sequential flow. The distinct types of data analysis supported by AIM are described in the following subsections in detail.

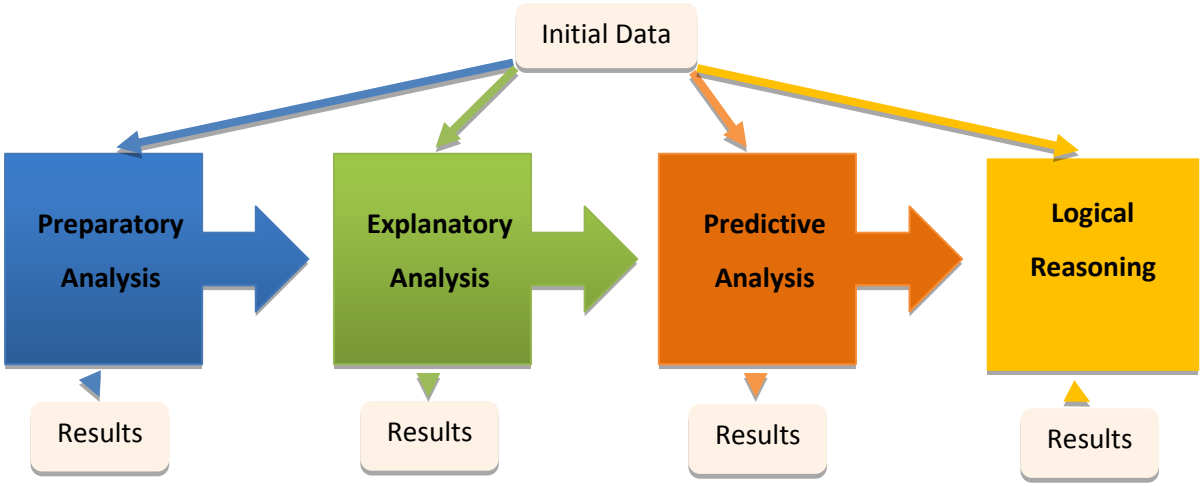


Figure 1. An Example Sequential Flow Integrating All Data Analysis Types.

1.1 Preparatory Analysis of Medical Data

Preparatory Analysis (or *Data Preparation*) is the process of pre-processing raw data to provide higher quality information for analysis. Essentially, it attempts to create more accurate, consistent and useful information from raw data by transforming them to new enhanced representations and addressing data abnormalities such as erroneous, irrelevant, conflicting, incomplete, redundant, insignificant and unrepresentative information. Preparatory analysis methods can be classified according to their specific aim in the following three categories, (a) Data Cleaning, (b) Feature Space Exploration and (c) Sample Space Exploration methods which are described below in detail. In the medical domain, numerous preparatory analysis frameworks have been developed to improve the quality of raw medical data collected from clinical practice [11-25].

1.1.1 Data Cleaning Methods

Data Cleaning (or *Data Cleansing*) is the process of analyzing raw data to correct erroneous, irrelevant, conflicting and incomplete information in order to improve data quality for further analysis. In the medical domain, erroneous patient information (e.g. out of range values in clinical measurements) may be generated from the manual data entry in clinical facilities. Irrelevant and conflicting medical information may be generated from integrating patient data from different medical departments or healthcare centers. Finally, incomplete medical information may be the result of irregular patient clinical visits.

These problems in raw data led to the development of data-cleaning methodological frameworks for various domains (e.g. epidemiological [14], genetic [15, 16] and clinical [17] data-cleaning frameworks). In the specific case of incomplete medical information, sophisticated data-cleaning methodologies are used for imputation of missing values, including the Multiple Imputation [18, 26], MissForest [19, 27] and Mixed-Effect Regression (or Mixed Model) [28, 29], with the latter method being especially appropriate for longitudinal data and generally for repeated correlated measurements of the same subjects. Data cleaning is the first important pre-processing phase of medical information that focuses on achieving an adequate level of data quality before further analysis is performed.

1.1.2 Feature Space Exploration Methods

Feature Space Exploration is the analysis process of subjects' characteristics to create a feature set with new enhanced characteristics and to address those that are redundant or insignificant, in the context of a specific data analysis problem. Feature (or variable) refers to a measurable characteristic (or attribute) of subjects. Feature space exploration methods can be classified in the following three categories, (a) Feature Transformation, (b) Feature Expansion and (c) Feature Selection methods which are described below in detail.

Feature Transformation (also called *Feature Extraction*) is the process of analyzing an initial set of features and subsequently replacing them with new enhanced features that are derived from the application of mathematical functions and algorithmic processes in the initial set. After feature transformation, the features may be completely different from the initial set and the feature space (in terms of features number) may increase, decrease or remain the same. In the context of medical data analysis, feature transformation can assist in the conversion of raw medical data to potentially more useful representations [20, 30]. For example, in case of multivariable depression analysis and given features weight (kg) and height (cm), a feature transformation method can be used to replace them with the body mass index ($index = \frac{weight}{height^2}$) due to literature association findings of the index with depression [31]. Feature transformation methods can also assist to meet assumptions of models that perform data analysis or improve their performance. For instance, in linear regression models, a log-transformation method used on the outcome variable may assist to improve model performance or meet the normality assumption of residual errors. Additionally, feature transformation methods such as variable discretization can improve significantly the analysis results (e.g. specific epidemiological analyses may be improved by discretization of patient

diagnosis year in intervals that correspond to different clinical protocol trends). Moreover, feature transformation methods such as is principal component analysis (PCA) can address possible correlations between features (by converting the initial feature set into linearly uncorrelated features, at the expense of interpretability) which may affect negatively specific data analysis models [21, 32, 33]. Finally, feature transformation methods include clustering techniques (that provide a cluster-membership feature based on an initial feature set representing meaningful subject groups) that may be useful in association, predictive and high dimensionality (large number of features) data analyses [22, 34, 35].

Feature Expansion (also called *Feature Construction)* is the process of analyzing an initial feature set to expand it with new features, utilizing mathematical functions and algorithmic processes. After feature expansion, the initial features remain the same and new features may be added in the initial set, thus feature space (in terms of features number) may increase. In the context of medical data analysis, feature expansion can provide new knowledge and insights to address medical research problems. For example, in medical analyses that evaluate the effect of treatment in disease outcomes where patients can receive different treatments (features) and combinations of two treatments, a feature expansion method that constructs and includes features of two-way treatment interactions[1] can be useful to assess the effect of treatment combinations over and above the additive effect of each single treatment (initial features) [36, 37].

Feature Selection (also called *Feature Reduction*) is the process of analyzing an initial feature set to select important features and remove those redundant and insignificant, utilizing mathematical functions and algorithmic processes. After feature selection, some of the initial features may not be in the final feature set and thus feature space (in terms of features number) may decrease. Feature selection methods can be divided in the following three categories, (a) Embedded methods, (b) Wrapper methods and (b) Filter methods. Hybrid implementations of the aforementioned categories can also be developed [38]. The first category, embedded methods, includes the most sophisticated data analysis models which select important features natively during their construction phase (e.g. LASSO regression [23, 39, 40]). The second category, wrapper methods, include data analysis models that do not

[1] An interaction exists between two features if the effect of one of them is different depending on the level of the other. A two-way (or first order) interaction involves two features that interact and it is constructed by their multiplication (higher order interactions involve larger number of features).

implement a feature selection process directly but are executed in many feature subsets (thus more computationally intensive than embedded methods) to derive the one with the highest model efficiency. Various wrapper methods have been developed that search feature space, determine feature subsets and compare their effect on model efficiency (e.g. Forward, Backward and Bidirectional/Floating Feature Selection methods [41] using regression or classification models). Statistical tests can be used to compare model efficiency between the feature subsets using appropriate measures according to the analysis type and target (e.g. in classification analyses the Correct Classification Rate measure and in regression analyses the Mean Absolute Error [42] and the Bayesian Information Criterion [43, 44] measures). Finally, the third category of filter methods focuses on the derivation of a feature ranking representing their importance that guides the selection process, however it also ignores entirely the overall model efficiency. Feature ranking is performed either with univariable measures (e.g. the Mutual Information Criterion, named also Information Gain [45, 46] and in regression problems the Pearson Correlation Coefficient [47]) or with multivariable measures (e.g. the Standardized Coefficients of regression models [48]) of feature importance. In the context of medical data analysis, feature selection methods have been applied in various domains to reduce the feature space by removing redundant and insignificant features [49].

Feature space exploration that includes feature transformation, expansion and selection methods is an important pre-processing phase of medical information that focuses on increasing the usefulness of data before further analysis is performed.

1.1.3 Sample Space Exploration Methods

Sample Space Exploration is the analysis process of the set of samples to derive new sample subsets or supersets that will address unrepresentative information by removing or duplicating samples. These new sample sets will be used for more accurate further analyses. Some sample space exploration methods are described below in detail.

Outlier Removal is the method of removing samples that are unrepresentative (outlier samples) and may affect the analysis towards inaccurate results. Outlier samples may be the result of errors in sampling process or variability in samples measurements, however in either case they should be addressed in order to avoid unrepresentative analysis results. In the context of medical data analysis, outlier removal has been applied in various domains to address unrepresentative information [24, 50-52].

Data Balancing is the method of removing (undersampling) or resampling (oversampling) samples randomly, in order to address unbalanced information in the analysis outcome that may lead to inaccurate results. For example, an analysis outcome may focus on the differences between male and female gender comparing them in a variety of characteristics. However, the gender target variable may contain unbalanced information with 10% females and 90% males. In this example, data balancing can resolve the target unbalance by either selecting randomly a subset of male samples with equal size to females (undersampling) or by resampling randomly female samples until the sample size for males is reached (oversampling). In the context of medical data analysis, data balancing has been applied in various domains [25, 53-55] to address unbalanced information than can affect statistical models' (such as linear and logistic regression models) towards inaccurate results.

Sample space exploration is an important pre-processing phase of medical information that focuses on increasing the robustness of data analysis to derive more accurate and consistent results.

1.2 Association Analysis of Medical Data

Association analysis in the statistical context, is the process of determining statistical differences, correlations and explanatory relationships in subjects' data. Statistical association analysis evaluates the association of one (univariate) or more (multivariate) target characteristics of subjects, with one (univariable) or more (multivariable) subject characteristics. This thesis includes univariable and multivariable association analyses with one target variable (univariate). In this section we describe the following three types of association analysis used in statistical modeling (a) Association Analysis of Difference, (b) Association Analysis of Correlation and (c) Explanatory Association Analysis.

Association Analysis of Difference is the process of identifying significant difference between two (or more) groups of subjects (represented by one input categorical variable) with regard to a target characteristic (target variable) of subjects. Statistical methods for univariable (one input categorical variable represents subject groups) association analysis of difference are provided in Table 1 (based on [56, 57]). Methods in Table 1 are firstly divided based on whether the compared groups contain correlated measurements (matched pairs). For example, groups contain correlated measurements if they observe the same subjects in different time points. Secondly, methods are divided based on the type of the measured target variable and its distribution type (continuous type and normal distribution, continuous or

ordinal type and not normal distribution, categorical type and time-to-event type). Thirdly, methods are divided based on the number of groups that are compared.

Table 1. Methods for Univariable Association Analysis of Difference.

Correlated Measurements *	Target Variable	Input Variable Categorical	Method
No	Continuous (Normal)	2 Groups	T-test
No	Continuous (Normal)	> 2 Groups	One-Way ANOVA F-test
No	Continuous/Ordinal (Non-normal)	2 Groups	Mann-Whitney U Test
No	Continuous/Ordinal (Non-normal)	> 2 Groups	Kruskal Wallis H Test
No	Categorical	2 Groups	Fisher's Exact Test (< 20 samples)
No	Categorical	> 2 Groups	Chi-square Test (≥ 20 samples)
No	Time to Event	> 2 Groups	Kaplan Meier Log-rank
Yes	Continuous (Normal)	2 Groups	T-test Matched Pairs
Yes	Continuous (Normal)	> 2 Groups	Mixed-effect Regression
Yes	Continuous/Ordinal (Non-normal)	2 Groups	Wilcoxon Signed-rank
Yes	Continuous/Ordinal (Non-normal)	> 2 Groups	Friedman Test
Yes	Categorical	2 Groups	McNemar's Test

* For example, repeated measurements of the same subjects between subject groups.

Association Analysis of Correlation is the process of identifying significant correlation between one characteristic of subjects (input variable) with regard to another target characteristic (target variable). Statistical methods for univariable (one input variable) association analysis of correlation are provided in Table 2 (based on [56, 57]) based on the type of input and target variables. Pearson correlation coefficient (r) is suggested for continuous input and target variables, Spearman rank correlation coefficient (r_s) when one or both (from input and target) variables are ordinal and Cohen's kappa correlation for categorical (input and target) variables.

Table 2. Methods for Univariable Association Analysis of Correlation.

Input and Target Variables	Method
Continuous (Normal)	Pearson Correlation
Continuous/Ordinal (Non-normal)	Spearman Rank Correlation
Categorical	Cohen's Kappa Correlation

Explanatory Association Analysis is the process of determining the association of one target characteristic of subjects (target variables) with one (or more) explanatory subject characteristics (input variables). Methods for multivariable (many input variables) explanatory association analysis are the statistical models provided in Table 3 (based on [56, 57]). Models in Table 3 are firstly divided based on whether the target variable contains correlated measurements such as repeated measurements of the same subjects. Secondly, models are divided based on the type of the measured target variable and its distribution type (continuous, ordinal, categorical and time-to-event).

Table 3. Models for Explanatory Multivariable Association Analysis.

Correlated Measurements *	Target Variable	Models
No	Continuous (Normal or non-Normal Residuals)	Linear Regression (with OLS or MLE, respectively)
No	Ordinal	Ordered Logistic Regression
No	Categorical (2 Values)	Binary Logistic Regression
No	Categorical (n Values)	Multinomial Logistic Regression
No	Time to Event	Cox Proportional Hazard Regression
Yes	Continuous (Normal or non-Normal Residuals)	Linear Mixed-Effect Regression (with OLS or MLE, respectively)
Yes	Ordinal	Generalized Estimation Equation
Yes	Categorical	Generalized Estimation Equation

* For example, repeated measurements of the same subjects between subject groups.

Statistical association analysis is the standard process of analyzing medical data to unravel differences, correlations and explanatory relationships in subjects' data [58].

1.3 Predictive Analysis of Medical Data

Predictive analysis, is the process of deriving predictions about unknown data from the application of advanced algorithmic processes in present and past information with an estimation of the process uncertainty. Predictive analysis is usually based on sophisticated machine learning models.

1.3.1 Machine Learning Models

Machine learning (ML) [59] is an evolving subfield of artificial intelligence that includes algorithmic processes able to provide predictions and improve their results through experience. Machine learning models range from statistical explanatory association models such as the Binary Logistic Regression model [60], which are highly interpretable from a mathematical perspective, to complex algorithmic processes such as the Random Forests [61, 62] and the Deep Neural Networks [63, 64] models which have a lower interpretability level.

Numerous machine learning models have been proposed in the medical domain to assist clinicians in disease diagnosis and prognosis phase [9, 65-69]. Predictive capabilities of the models in the evolving subfield of machine learning in the artificial intelligence domain is a powerful state-of-art approach to enhance the clinical decision-making process in the medical domain.

1.3.2 Model Performance Evaluation

Model performance evaluation is an important concept in machine learning modeling that focuses on the estimation of accuracy and uncertainty in the predictive process while reducing possible bias. A popular sophisticated process for ML model performance evaluation is the cross-validation [70]. The essence of model performance evaluation is that models should be evaluated in independent datasets that were not used in model construction phase in order to minimize bias. Other simpler approaches for model performance evaluation have also been proposed [71].

1.4 Logical Analysis (Reasoning) on Medical Data

Logical analysis or logical reasoning, is the inference process of new knowledge based on existing *facts* and conditional statements defined as *rules*. A fact is a conclusion that can be inferred without any required preconditions. A rule defines a conditional statement where a set of preconditions should be satisfied for a specific conclusion to be inferred. Thus, a fact can be considered as a special type of rule with an empty set of preconditions. In logical reasoning, rules can be chained. A rule is *chained* to another rule if a conclusion of the first is precondition to the second. Thus, the inference process of a rule can trigger the inference processes of its chained rules.

Basic forms of logical reasoning are the deductive, inductive, abductive and defeasible reasoning. Logical reasoning forms are classified into monotonic and non-monotonic reasoning categories described in the following section.

Logical reasoning can be applied to simulate the human logic for decision-making [72]. The present study focuses on the logical reasoning form of defeasible reasoning that is able to simulate clinicians' decision-making processes that may involve uncertain and contradictory medical evidence.

1.4.1 Monotonic and Non-monotonic Reasoning Categories

Logical reasoning forms are classified into monotonic and non-monotonic reasoning [73] categories. The basic principle that distinguishes monotonic from non-monotonic reasoning is that the former supports information that once derived it cannot be invalidated while the latter supports also information that can be invalidated with future knowledge under certain conditions.

Monotonic reasoning is logical reasoning where the addition of new knowledge (preconditions) cannot invalidate (decrease) previous knowledge that was derived from the rules. For example, global truth rules such as "if the planet is earth (precondition) then (rule) it's shape is round (conclusion)" are a type of monotonic reasoning.

Non-monotonic reasoning is logical reasoning that can derive information which in the presence of future knowledge and under certain conditions can be invalidated. An example of non-monotonic reasoning is the following, given the fact that patient A is diagnosed with rheumatoid arthritis (RA) and a rule suggesting treatment B for patients diagnosed with RA, then we should derive that patient A should receive this treatment. However, future knowledge that treatment B was not effective on patient A should invalidate the conclusion that patient A should receive treatment B. Non-monotonic reasoning can

address scenarios of uncertainty and contradicting evidence where already derived information may be invalidated by future knowledge.

1.4.2 Deductive, Inductive, Abductive and Defeasible Reasoning

Deductive, inductive, abductive [72, 74] and defeasible reasoning [75] are different forms of logical reasoning. The basic difference that distinguishes them is the logical concepts derived from the reasoning process. In deductive reasoning the logical concepts derived from the reasoning process are conclusions while in inductive and abductive reasoning they are rules and preconditions respectively. Defeasible reasoning is a special type of non-monotonic reasoning that extends the ordinary deductive reasoning with defeasible conclusions.

Deductive (*top-down*) *reasoning* is the process of extracting conclusions, given a set of preconditions and rules that guarantee the truth of the conclusions given the truth of the preconditions. Ordinary deductive reasoning is monotonic in the sense that a derived conclusion cannot be withdrawn when knowledge is increased. An example of deductive reasoning is the following, given the fact that patient A has fever for many days and a rule specifying that if a patient has fever for many days (precondition) then he should receive medication (conclusion), we can derive that patient A should receive medication. Additionally, another example of a deductive reasoning process in the computational modeling context, can be considered the execution of a well-defined predictive model to extract predictive information.

Inductive (*bottom-up*) *reasoning* is the process of extracting generalization rules, given numerous observations of preconditions and conclusions. For example, given a set of facts that lists employees, their productivity level and disease status where the majority that have the flu also have low productivity, we can derive the generalization rule that the flu probably leads to low productivity. In this type of reasoning, rules do not guarantee the truth of the conclusions given the truth of the preconditions, since conclusions are not certain but probable and thus, it inductive reasoning belongs to non-monotonic reasoning category. Additionally, another example of an inductive reasoning process in the computational modeling context, can be considered the configuration of variable parameters in the construction process of an explanatory or predictive model.

Abductive reasoning is a special type of reasoning with incomplete knowledge in the preconditions and rules. It is an extension of deductive reasoning that is used to derive the most probable explanation (evidence) given a set of observations. For example, given a set of

rules that specify medical conditions (preconditions) leading to certain symptoms (conclusions), then for a specific patient's symptoms we can derive the medical condition that explains most of them. Abductive reasoning implements the process of extracting the most probable set of preconditions that explains a conclusion (s), given a set of conditional statements (rules). Thus, due its probabilistic nature, it does not guarantee the truth of the conclusions, similarly to inductive reasoning and it belongs to non-monotonic reasoning category. Additionally, another example of an abductive reasoning in the computational modeling context as an abstract process, can be considered the development of a hypothesis or the development of a model's structure (configuration of variables) that are associated with a target characteristic.

Defeasible reasoning is a special type of non-monotonic reasoning extending the ordinary deductive reasoning with defeasible conclusions. Defeasible conclusions are conclusions that may be invalidated by future knowledge. In this type of reasoning, rules do not guarantee the truth of the conclusions given the truth of the preconditions, since conclusions may be withdrawn in the future. Defeasible reasoning can address scenarios of uncertainty and contradicting evidence. For example, a scenario that includes uncertainty from incomplete evidence is the following, a clinical policy rule may define that "if a patient is in pain (precondition 1) and there is no information on patient's allergic reaction with drug A (precondition 2) then he should receive therapy with drug A (conclusion)". The second precondition denotes the absence of information on patient's allergic reaction with drug A, but still it can derive a useful conclusion which in the light of future knowledge (on patient's allergic reactions with drug A) it will be invalidated. Another example that includes contradicting evidence is the following, a clinical rule may define that "if the patient has pain (precondition) then he should receive pain-related therapy with drug A (conclusion)" while a second more important rule may define that "if the patient is allergic to drug A (precondition) then he should not receive therapy with drug A (conclusion)". In this example, the two rules have contradicting conclusions while the fact that the second rule is more important to the first denotes a policy to resolve conflicts that may arise from patients that satisfy the preconditions of both (having pain and allergy to drug A). Defeasible reasoning can support the aforementioned scenarios that include uncertain and contradicting evidence using defeasible conclusions, conflict resolution policies and preconditions that denote information absence.

1.4.3 Defeasible Logic

Defeasible Logic is an efficient rule based non-monotonic defeasible reasoning formalism that is able to derive conclusions from incomplete and conflicting information [76, 77]. The basic concepts of defeasible logic are described below, adjusted for the medical domain with relevant examples. The uppercase words denote variables in the examples.

- *Facts:* Facts are rules that derive conclusions without preconditions and cannot be invalidated. In essence they express indisputable statements. For example, a fact can be related to the following knowledge, "*The patient with unique identifier (id) 12345 is diagnosed with rheumatoid arthritis and this fact cannot be contradicted by any other medical evidence.*". This knowledge can be represented in Defeasible Logic with the following rule labeled as "*rule_1*". Note the conditional statement symbol → of rule labeled "*rule_1*".

  $\boldsymbol{rule\_1}$: $\rightarrow disease\_RA\_positive\_diagnosis\_for\_patient\_id\ (12345)$

- *Strict Rules:* Strict rules are rules that have preconditions and derive conclusions which cannot be invalidated if the preconditions hold. For example, a strict rule can be related to the following knowledge, "*If a clinician assessed that a patient (specified by variable ID unique identifier) has RA (rheumatoid arthritis) disease then it is concluded that this patient is diagnosed positive for RA disease and this conclusion cannot be contradicted by any other medical evidence*". This knowledge can be represented in Defeasible Logic with the following rule labeled as "*rule_2*". Note the conditional statement symbol → of rule labeled "*rule_2*".

  $\boldsymbol{rule\_2}$: $clinician\_assessed\_disease\_RA\_positive\_for\_patient\_id(ID)$
  $\rightarrow disease\_RA\_positive\_diagnosis\_for\_patient\_id\ (ID)$

- *Defeasible Rules:* Defeasible rules are rules that have preconditions and derive conclusions which can be invalidated by contradictory conclusions of other rules. For example, a defeasible rule can be the following "*If a patient (specified by variable ID unique identifier) has positive results for RA diagnosis based on a low-specificity exam then it is concluded that the patient is diagnosed positive for RA until stronger contradictory medical evidence is provided.*". This knowledge can be represented in Defeasible Logic with the following rule labeled as "*rule_3*". In this example, we have not yet discussed rules of stronger contradictory evidence. Note the conditional statement

symbol $\Rightarrow$ that specifies the defeasible conclusion which can be invalidated by contradictory conclusions of stronger rules.

$\boldsymbol{rule\_3}$: $low\_specificity\_exam\_is\_RA\_positive\_for\_patient\_id(ID)$
$\Rightarrow disease\_RA\_positive\_diagnosis\_for\_patient\_id\ (ID)$

- *Priority Rule* or *Superiority Relation:* The priority rule is a rule that defines a binary acyclic relation over a pair of rules with contradictory conclusions to specify which rule is stronger that thus its conclusion should prevail. An example of knowledge related to contradictory conclusions of defeasible rules and priority rules among them is the following "*If a patient (specified by variable ID unique identifier) has positive results for RA diagnosis based on a low-specificity exam then it should be concluded that the patient is diagnosed positive for RA until stronger contradictory evidence is provided. However, if a clinician assessed that the aforementioned patient does <u>not</u> have RA disease then it should be concluded that he is <u>not</u> diagnosed positive for RA since the clinician assessment provides stronger medical evidence than the low-specificity exam*". This knowledge can be represented in Defeasible Logic with the following rules.

$\boldsymbol{rule\_4}$: $low\_specificity\_exam\_is\_RA\_positive\_for\_patient\_id(ID)$
$\Rightarrow disease\_RA\_positive\_diagnosis\_for\_patient\_id\ (ID)$

$\boldsymbol{rule\_5}$: $clinician\_assessed\_disease\_RA\_positive\_for\_patient\_id(ID)$
$\Rightarrow \neg\, disease\_RA\_positive\_diagnosis\_for\_patient\_id\ (ID)$

$\boldsymbol{rule\_6}$: $\rightarrow superior(rule\_5, rule\_4)$

Defeasible rules labeled as "*rule_4*" and "*rule_5*" can lead to a knowledge conflict if both their preconditions hold for the same patient since the former concludes that a patient is diagnosed positive for RA while the latter concludes that a patient is <u>not</u> diagnosed positive for RA. Note the negation symbol ¬ in front of the conclusion "*disease_RA_positive_diagnosis_for_patient_id(ID)*". The priority rule labeled as "*rule_6*" can resolves this conflict by specifying that rule labeled "*rule_5*" is stronger (superior) than rule labeled "*rule_4*" and thus its conclusion prevails.

- *Conflicting Literals:* Conflicting Literals essentially express constraint rules that define a set of literals from which any selected pair when derived from theory, leads to a conflict

of contradictory knowledge. An example of knowledge related to Conflicting Literals is the following, "The RA disease level of a patient (specified by variable ID unique identifier) can fall into only one of the following three levels, *low*, *medium* and *high*. Evidence for more than one RA disease level for a specific patient leads to contradicting conclusions.". This knowledge can be represented in Defeasible Logic with the following rule labeled as "*rule_7*".

$$\boldsymbol{rule\_7}: \quad \rightarrow conflicting\_pairs :: disease\_RA\_level\_low(ID), disease\_RA\_level\_medium(ID),\ disease\_RA\_level\_high(ID)$$

- *Defeaters:* Defeater rules are rules that derive conclusions if two requirements are fulfilled, (a) their preconditions hold and (b) there are contradictory conclusions derived from other theory rules. In essence, their only difference to defeasible rules is that there should be a contradictory conclusion derived from the theory in order to be triggered. They only exist to invalidate contradictory evidence and thus without contradictory evidence they don't provide any information. Apart from this difference they behave similarly to defeasible rules and are specified by symbol $\rightsquigarrow$ of conditional statement.

Defeasible logic is a "skeptical" logic since there might be evidence for conclusion A but also evidence for the contradictory conclusion ¬A (negation of A) and in this conflicting case neither conclusion is derived but the logic consults the rules' priority relations to determine which rule has higher priority over the others in order to resolve the conflict. If the evidence for A has higher priority over the evidence for ¬A then A is concluded.

In defeasible logic, evidence that is provided for a conclusion from a set of rules can be categorized into *definitely provable* and *defeasibly provable* which are described below in detail. Definitely provable are the conclusions that can be derived using only facts and strict rules. More specifically a conclusion is definitely provable in the three following cases:

- If it is a fact.
- If it is a conclusion of a strict rule that its preconditions are satisfied.
- If it is a conclusion of a strict rule that can be derived based only on strict rules and facts.

Defeasibly provable conclusions are a superset of definitely provable conclusions that include also conclusions which require at least one defeasible rule in order to be derived. More specifically a conclusion is defeasibly provable in the three following cases:

- If it is definitely provable.
- If it is a conclusion of strict or defeasible rule that its preconditions are satisfied and all contradictory rules are not satisfied.
- If it is a conclusion of a strict or defeasible rule that is satisfied and it is stronger from all contradictory rules based on priority relations.

Defeasible reasoning frameworks such as Defeasible Logic are able to support natively incomplete and conflicting information. Thus, they can be used to formalize and simulate the medical reasoning process which may include uncertain and contradictory medical evidence. AIM (artificial intelligence in medicine) technology should support defeasible reasoning capabilities to simulate clinical decision-making processes of many clinical protocols, information uncertainty and knowledge conflicts. Defeasible reasoning can also assist in incorporating medical research results (from explanatory and predictive analyses) in the decision-making process that may also create conflicting information.

Defeasible reasoning can thus provide a unified framework to support simulation of medical reasoning logics both from the clinical practice and medical research domains, able to resolve any arising knowledge conflicts.

CHAPTER II. Artificial Intelligence in Rheumatoid Arthritis

Artificial intelligence in medicine (AIM), is state-of-art information technology that enables the collection and analysis of medical information and the implementation of appropriate actions, to support disease management, therapy, diagnosis, prognosis and prevention strategies. Sophisticated AIM frameworks can incorporate preparatory, explanatory and predictive data analysis methods and logical reasoning. These frameworks can unravel data patterns and associations, prognostic and predictive insights, and are also able to simulate the medical reasoning process of clinical decision-making.

The present study focuses in AIM technology in the specific domain of rheumatoid arthritis long-term prognosis and disease management under therapy. The next subsections are structured as follows, the first subsection describes the basic aspects of rheumatoid arthritis (RA) and its therapy, the second refers to the capabilities of AIM technology in RA prognosis and management, while the third analyzes literature limitations in the RA long-term prognosis under therapy.

1. Rheumatoid Arthritis (RA), Disease Activity and Medication

Rheumatoid arthritis (RA) is a chronic autoimmune disease that causes inflammation to body joints and extra-articular inflammation. Symptoms range from fatigue, joint tenderness, joint swelling and pain to permanent joint damage, disability and reduced life expectancy, while the cause of the disease is still unknown. Several co-morbidities affect both quality of life and increase morbidity and mortality. Rheumatoid Arthritis affects 1% of the world population (1.29 million people in America and 2.9 million people in Europe). RA medications primarily focus on controlling the symptoms and slowing the progression of the joint damage in order to improve quality of life.

Disease activity state of rheumatoid arthritis is very important in RA patients' treatment course to assess their long-term prognosis [78]. According to the widely accepted treat-to-target (T2T) strategy, the aim of treatment in RA is to improve patients' health-related quality of life by abrogation of inflammatory burden [79]. T2T approach in RA treatment aims at targets of remission or low state of disease activity which is evaluated by evidence-based measures such as the disease activity score of 28 joint counts (DAS28) [80]. DAS28 score has been established as one of the indexes applicable both in clinical practice and in clinical trials to assess disease activity state and guide treatment modifications [81].

Apart from DAS28 score, other composite measures have also been developed for use in clinical practice such as the CDAI and SDAI indexes [82].

Medications for Rheumatoid Arthritis are categorized to drugs that help to ease the symptoms such as Nonsteroidal Anti-inflammatory Drugs (NSAIDs), and drugs that aim to slow the disease progression such as Corticosteroids, Conventional Synthetic Disease Modifying Anti-Rheumatic drugs (csDMARDs), Biologic Disease Modifying Anti-Rheumatic drugs (biologic DMARDs), and recently Janus Kinase Inhibitors (JAK inhibitors). Each of these treatments differs in their mechanism of action. A different combination and sequence of treatments maybe required for each patient. The present study focuses on RA treatment with biologic DMARDs.

Biologic DMARDs (bDMARDs) are potent agents that control inflammatory burden of RA, leading to improved prognosis of the patients. Clinical effectiveness of biologic agents has been extensively reported in controlled clinical trials and registries. Different registries have shown that patients' treatment with biologic anti-TNF (tumor necrosis factor alpha inhibition) agents induce remission at 12 months in 15-35% of patients [83-86] and low disease activity in 27-54% of patients, while figures are rather similar with biologic non anti-TNF agents such as Tocilizumab (36% remission and 50% low disease activity at 12 months) [87].

2. Artificial Intelligence in RA Management and Prognosis

Artificial intelligence in medicine (AIM), is state-of-art information technology that has the potential to improve rheumatoid arthritis management and prognosis towards better therapy outcomes. AIM technology capabilities for rheumatoid arthritis management and prognosis, are described below.

AIM technology includes electronic health record systems (EHRs) that can support rheumatoid arthritis information recording and management, such as patient's medical history, symptoms, diagnosis and treatments [3, 88]. EHRs may provide services both for healthcare providers and for patients (PHRs). PHRs may include services for patient self-health monitoring focused on rheumatoid arthritis, such as health assessment questionnaires (e.g. HAQ [89]), recording of daily life observations and diet calendars which empower patients to take more active role in their own health management [4, 90]. In addition, AIM technology is able to improve communication flows between patients and clinicians during

rheumatoid arthritis therapy, using services such as appointment scheduling, reminders and messages that facilitate patient-clinician interaction [5].

AIM technology can also support rheumatoid arthritis prognosis by providing access to the results of medical research [91] while also enabling advanced data access control policies to manage information sharing and dissemination of medical knowledge [5]. In addition, AIM technology includes clinical decision-support systems (CDSs) that can support patient outcome predictions for rheumatoid arthritis [92, 93], intelligent alerts and medical recommendations [94]. AIM clinical decision-support systems are able to formalize and simulate clinicians' decision-making rules, unravel medical data insights and provide predictive information, to advance the patient therapy process. Overall, AIM technology has the potential to improve significantly rheumatoid arthritis management and prognosis.

3. Literature Limitations in RA Prognosis Under Biologics

Biologic Disease Modifying Anti-Rheumatic drugs (biologics or bDMARDs) include potent agents to control rheumatoid arthritis progression. However, approximately 30-50% of patients on biologic drug treatments, although they improve disease activity status, still have moderate disease activity (MDA) after the first treatment [85]. Moreover, moderate disease activity may be present in a significant part of their long-term biologic therapy despite of treatment drug switches. Essentially, this is a group of patients with medium disease under biologic treatments that neither improve nor deteriorate outside moderate disease activity for a significant time (persistent MDA).

Long-term outcomes for the persistent MDA group of RA patients who are treated on biologics but still have long-duration medium disease, are scant in the literature. First of all, the size of the persistent MDA group in rheumatoid arthritis clinical practice is unknown. It is an open medical research question whether they comprise the majority of RA patients under biologic therapy or a smaller proportion. In addition, these patients have exhibited long-duration medium disease under biologic therapy, however it is also unknown what is the rest of their disease course and whether their overall disease course is closer to persistent low or high disease. Finally, data are unavailable about the long-term outcomes of the persistent MDA group such as their long-term functionality status or the number of serious adverse events that occurred during therapy.

Literature studies that analyze long-term outcomes of RA patients with persistent MDA under treatment have mainly focused on early disease rather than established RA and

on conventional (csDMARDs) rather than biologic treatments (bDMARDs) [95, 96]. Another study compared outcomes of RA patients that achieved different disease levels, including MDA level, at 12 months of therapy with biologics (based on a pooled analysis of randomized controlled clinical trials in anti-TNF agents), nevertheless longer-term outcomes are unavailable in this study [78]. The literature research showed that long-term outcomes of rheumatoid arthritis patients with persistent MDA under biologics, are rather scant.

Part of the present study provides statistical analyses that compare the long-term outcomes of the following three RA patient groups, long-duration low/remission (persistent LDA), medium (persistent MDA) and high (persistent HDA) disease, under biologic therapy and irrespective of fluctuations. The analyses provide insights on persistent MDA patient group in clinical practice, presenting information on the size of this group and its long-term prognosis, which can assist in treatment tailoring to improve outcomes. All analyses conducted in this study are provided as services of an artificial intelligence clinical decision-support system framework developed for rheumatoid arthritis management and long-term prognosis. The next chapter describes the objectives, methodology and implementation results of the CDS-RA system framework, in detail.

CHAPTER III. Clinical Decision Support (CDS) System for RA

This chapter presents CDS-RA, a clinical decision-support system framework for (early and established) rheumatoid arthritis management and long-term prognosis. Initially, the CDS-RA framework aims to provide information on the long-term patient outcomes (regarding patient functionality and serious adverse events) associated with patient groups that exhibit persistent low (LDA), medium (MDA) and high (HDA) disease under biologic therapy and also analyze further the heterogeneity of patients with persistent medium disease. In addition, it aims to provide a complete prognostic system able to predict and reason about the persistent disease level that a patient may develop. The system has artificial intelligence characteristics and functionality that includes prognostic services, association analyses and data management services in order to assist the clinicians early, in the process of decision-making and personalized treatment towards improved outcomes.

This chapter is structured as follows. Section 1 presents the objectives of the CDS-RA system framework, section 2 analyzes the development methodology of the CDS-RA system framework and section 3 presents the system's implementation results.

1. Objectives

This section describes the core objectives of the CDS-RA system framework. The definition of persistent disease level is crucial in the objectives that the system attempts to achieve. *Persistent disease level* (*PDL*) in the present study was defined as the same disease activity level (DAS28 within a specific range), exhibited by a patient under biologic therapy, for at least half of his 5-year clinical follow-up, cumulatively and irrespective of fluctuations. Three PDL patient groups were specified, the LDA (DAS28 range low or remission), MDA (DAS28 range moderate) and HDA (DAS28 range high) groups, respectively. The analysis of a large RA patient cohort with long-term biologic therapy (at least 5 years) was required to achieve these aims. The distinct objectives of the CDS-RA framework are described below in detail:

- The first objective of the CDS-RA framework focuses on the long-term outcomes associated of the MDA group, since long-term information on this group (RA patients of persistent medium disease under biologic therapy) is limited. More specifically, the framework aims to utilize preparatory and association analyses to compare the long-term

(5-year) trajectory of functionality (HAQ) and the cumulative serious adverse events (SAEs) and functionality (HAQ) at 5 years between the MDA patient group and the LDA and HDA patient groups.

- The second objective of the framework of the CDS-RA framework focuses on the analysis of MDA group for internal heterogeneity in MDA patient subgroups, since this information can potentially assist clinicians in further targeting treatments for these subgroups, in order to improve outcomes. More specifically, the framework aims to utilize preparatory and association analyses to compare the long-term (5-year) trajectory of functionality (HAQ) and the cumulative serious adverse events (SAEs) and functionality (HAQ) at 5 years, between two distinct MDA patient subgroups, the *lower-MDA* (persistent lower medium disease) and the *higher-MDA* (persistent higher medium disease) subgroups.

- The third objective of the framework of the CDS-RA framework focuses on the early prediction of a patient's PDL group. More specifically, the framework aims to utilize predictive analysis based on patients' early therapy data (first 6 to 9 therapy months) in order to support personalized predictions regarding the development of persistent low (LDA) medium (MDA) or high (HDA) disease.

- The fourth objective of the CDS-RA framework focuses on the patient PDL (LDA, MDA or HDA) group prognosis, utilizing and prioritizing between many different medical evidence sources. More specifically, the framework aims to develop an artificial intelligence (AI) engine that will initially support three policies for patient PDL group prognosis based on the following three medical evidence sources in decreasing priority, (a) patient's long-term follow-up data, if they exist, which fulfill a group's classification criteria by definition, (b) clinicians' expert opinion and (c) the predictive analysis that utilizes patient's early (first 6 to 9 months) therapy data (third objective). The engine will support a rule-based AI logic for medical reasoning that will be accessible, reusable, configurable and extendable with additional prioritized medical policies (rule priorities resolve possible conflicts of contradicting medical evidence).

- The fifth objective of the CDS-RA framework focuses on supporting an easy-to-use, web-based, mobile-compatible platform for online data entry, visualization and analysis, targeted both for patient and clinician users. The framework aims to automate the patient-

clinician interaction during therapy by supporting frequent, cost-effective and convenient online data collection and analysis, compared to the traditional time-consuming, costly and inconvenient process of data collection during clinical visits, which can be conducted only in trimester and semester basis.

2. Methodology

The development methodology of the CDS-RA system framework is provided in the following sections. More specifically, section 2.1 describes the framework's supported functional and non-functional services, section 2.2 presents the framework's architectural design, section 2.3 provides information on the patient data sources that were analyzed and patient eligibility criteria for analysis, and finally section 2.4 describes the artificial intelligence methods for medical data analysis that were conducted for data preparation, data associations derivation, data predictions and logical inferences.

2.1 Functional and Non-Functional Requirements

The CDS-RA framework provides services both for patient and clinician users. Figure 2 illustrates the functional and non-functional requirements of the framework. Essentially, the system supports two user modules, the patient module and the clinician module.

The patient module provides the following services, (a) a self-assessment of the patient's pain level using visual analogue scale (VAS Pain), (b) a self-assessment of the patient's functionality using the health assessment questionnaire (HAQ) [89], (c) a self-assessment of the patient's quality of life using the EUROQOL questionnaire [97]. The traditional process of collecting this information in clinical practice is based on inconvenient, time-consuming and costly trimester and semester clinical visits. The system overcomes these barriers of patient data collection by providing the aforementioned services as web applications, accessible from a web-browser without any software installation requirements. The patient is able to access the online services conveniently from his home environment. In addition, the services are compatible with mobile devices focusing on tablet computers, in order to enable also outdoor usage. The patient data collected over time are stored into a central repository to be used by the clinician module.

The clinician module provides the following services:

- Service 1. Presentation of patient's demographics.
- Service 2. Presentation of patient's disease activity progression graphs.
- Service 3. Prognosis of the long-duration (persistent) disease level (PDL), low (or remission), medium or high, that the patient will probably develop, using logical rules that form the following three policies of decreasing priority.
  - Policy 1. Logical rules that derive patient's PDL based on any available long-term data on patient's longitudinal disease activity from clinical assessments (PDL inference by definition).
  - Policy 2. Logical rules that derive patient's PDL based on clinician's expert opinion (PDL inference by clinician user input).
  - Policy 3. Logical rules that derive patient's PDL based on patient's early therapy data from first 6-9 therapy months (PDL inference by predictive modeling).
- Service 4. Presentation of significant differences in patient longitudinal functionality (HAQ course) and serious adverse events (cumulative SAEs course) between RA patients grouped by persistent disease level (PDL), low (or remission), medium and high.
  - Association analysis of patients' persistent disease level (low, medium and high) with longitudinal functionality (HAQ), during 5 years of therapy in a large RA cohort.
  - Association analysis to compare patient cumulative serious adverse events (SAEs) and functionality (HAQ) at 5 years of therapy, between patient groups of different persistent disease levels (low, medium or high) in a large RA cohort.
- Service 5. Recording of the following clinical information from the clinician's assessment which will be linked to a specific patient's self-assessment:
  - Clinical parameters: Visual Analogue Scale for Physician Assessment (VAS Physician), Erythrocyte sedimentation rate (ESR), C-reactive protein (CRP).
  - Free-text clinical notes.
  - Selection of patient joints that are tender and swollen (TJC28: Tender joint count 28, SJC28: Swollen joint count 28) from an interactive image illustrating joints that can

be affected by rheumatoid arthritis. Alternatively, the clinician can add the sum of tender and swollen joints that are affected.

- Presentation of a summary table of the following 16 clinical parameters based on a specific patient self-assessment data and its corresponding clinical assessment data, Tender joint count 28 (TJC28), Swollen joint count 28 (SJC28), Tender joint count 44 (TJC44), Swollen joint count 44 (SJC44), Visual analogue scale for physician assessment in range [0-100] (VAS Physician), Visual analogue scale for patient's pain self-assessment in range [0-100] (VAS Pain), Visual analogue scale for patient's global self-assessment in range [0-100] (VAS G), Health assessment questionnaire score (HAQ), Erythrocyte sedimentation rate in mm/h measurement unit (ESR), C-reactive protein in mg/L measurement unit (CRP), Disease activity score of 28 joint counts with ESR and 4 variables (DAS28-ESR) [80], Disease activity score of 28 joint counts with ESR and 3 variables (DAS28-ESR(3)) [80], Disease activity score of 28 joint counts with CRP and 4 variables (DAS28-CRP) [98], Disease activity score of 28 joint counts with CRP and 3 variables (DAS28-CRP(3)), Simple disease activity index for rheumatoid arthritis (SDAI) [99], Clinical disease activity index for rheumatoid arthritis (CDAI) [100]. Disease activity level semantics (remission, low, moderate and high) will be provided for DAS28, SDAI and CDAI scores.

The clinician module also supports the non-functional requirements of web accessibility and mobile compatibility such as the patient module, focusing on tablet computers, to enable ease-of-use and enhance user adoption.

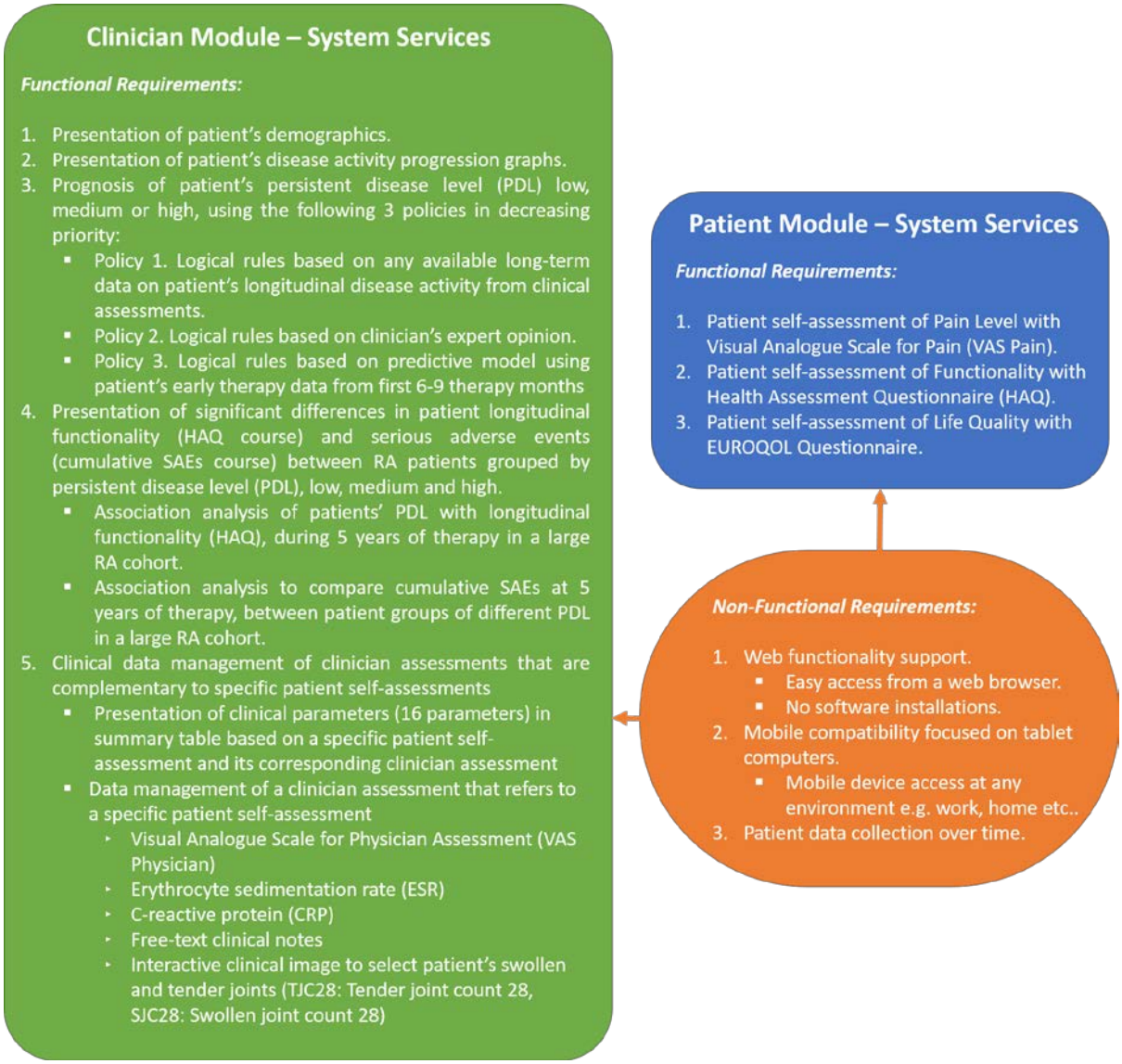


Figure 2. Functional and Non-Functional Requirements of Clinician and Patient Modules.

2.2 Architectural Design

The architecture of the CDS-RA framework consists of the following three layers, A. the Data Storage, B. the Artificial Intelligence and C. the End-User Layer. Figure 3 illustrates the architectural design of the CDS-RA framework in detail.

Essentially, the Data Storage Layer depicts the process of constructing the main patient database of CDS-RA framework. Rheumatoid arthritis patient data sources from seven hospital facilities across Greece (PAGNI at Crete, NIMITS at Athens, Euroclinic at Athens, Laikon at Athens, Sismanoglio at Athens, PGNI at Ioannina and General Hospital Kavalas) are integrated in PAGNI hospital registry database. Extract, transform and load (ETL) operations are performed on PAGNI hospital registry database and data from a

selected RA patient cohort treated on 5-year biologic therapy, are loaded into a dedicated RA database which consists the main database of the CDS-RA framework.

The Artificial Intelligence Layer illustrates all the statistical analyses, predictive modeling and logical reasoning processes that are conducted to provide results that are presented to the clinician to support clinical decision-making. Initially, a preparatory analysis is conducted to preprocess the patients' data from CDS-RA database. In this analysis missing longitudinal patient data are imputed and summarized in clinically significant intervals. The preprocessed patient data are used in association and predictive analyses. Two association analyses are performed after grouping patients based on their persistent disease level (PDL) low (low/remission), medium, or high. The first focuses on the association of the groups with distinct 5-year functionality (HAQ) courses and the second on the comparison of groups' cumulative serious adverse events at 5 years. A predictive analysis is also performed to construct a model able to predict a patient's persistent disease level (PDL) based on early patient therapy data. The performance of the model is evaluated in a separate process. The last analysis included in this layer is the *AI Medical Reasoning Logic* which essentially implements logical reasoning to derive a prognosis of the patient's persistent disease level (PDL) from multiple prioritized policies that use different sources of medical evidence with different reliability level.

Finally, the End-User Layer contains all the web services and user interfaces of the CDS-RA system framework. The web services utilize the results from the analyses conducted in the Artificial Intelligence Layer and also patient data (extract and load operations) from the main database of the CDS-RA framework, in order to provide the concrete functionality that was described in the previous section in detail.

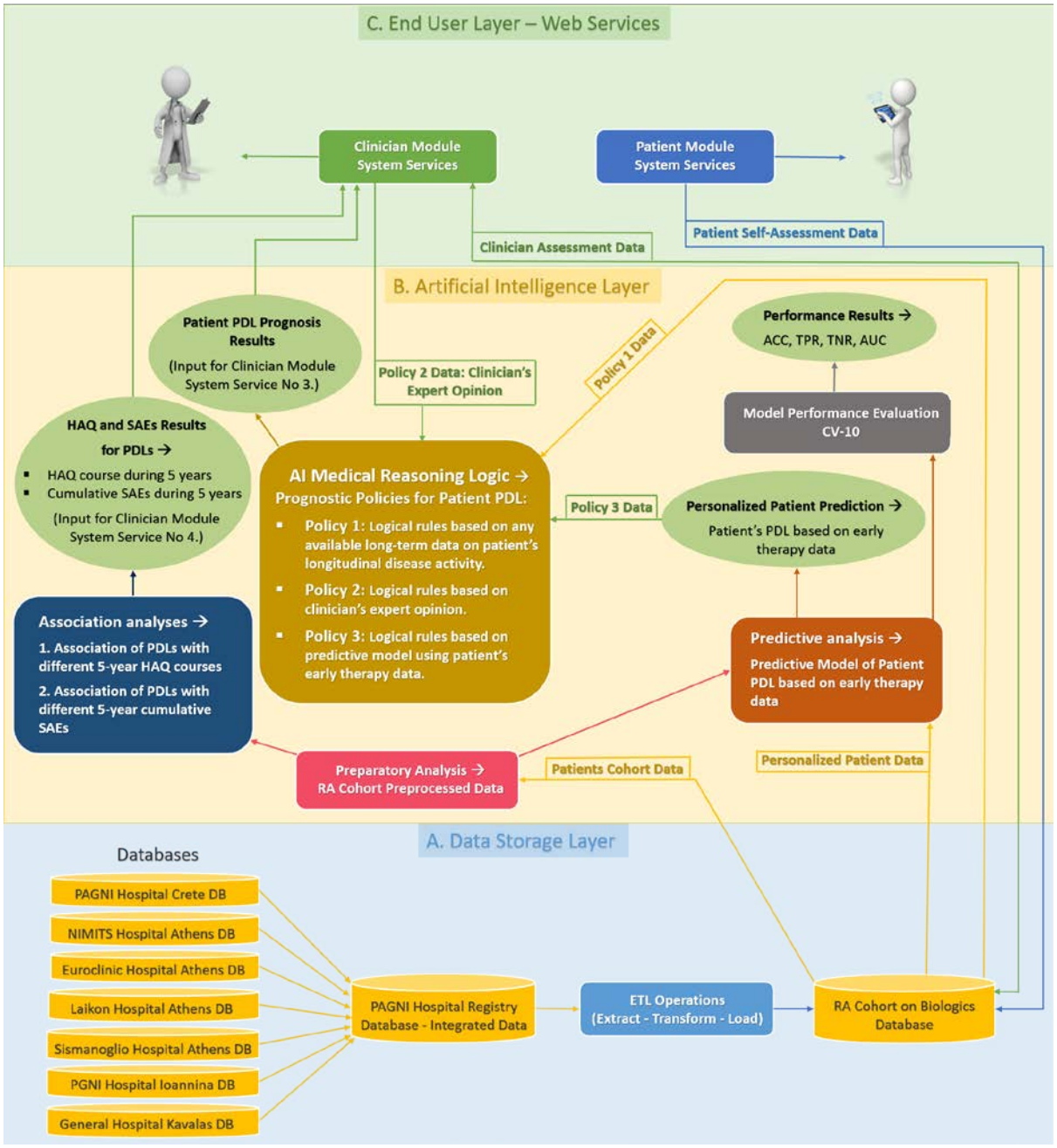


Figure 3. The architectural design of the CDS-RA system framework.

2.2.1 Technological Frameworks

The development of the CDS-RA architectural infrastructure (Figure 3) was based on the following technologies:

- The CDS-RA system main database "RA Cohort on Biologics" in the Data Storage Layer was developed in MySQL relational database.

- The CDS-RA web services and end-user functionality (in the End-User Layer) was developed in Java (programming language) [101] and Spring MVC (Java web application framework) [102] which were used as back-end technologies and in HTML 5 [103], JSP [104] with JSTL [105] Views, Twitter Bootstrap [106] which were used as front-end technologies. The back-end technologies were also used in the integration of all the architectural layers and in the development of ETL operations (Data Storage Layer).
- Data preparatory analysis (in the Artificial Intelligence Layer) to support longitudinal data summarization was implemented in Java and Spring MVC.
- Data preparatory analysis (in the Artificial Intelligence Layer) to support missing longitudinal data imputation was implemented in MATLAB [107].
- The association and predictive analyses (in the Artificial Intelligence Layer) were implemented in MATLAB.
- The Logical Analysis (in the Artificial Intelligence Layer) (in the Artificial Intelligence Layer) was developed in Prolog [108] using XSB Engine (https://xsb.com/xsb-prolog) [109] with the Defeasible Logic extension using Defeasible Logic Metaprogram (Appendix C. and [110]).

2.3 Patients Data Sources and Eligibility Criteria

This section describes the patients included in the PAGNI Hospital Registry Database and the eligibility criteria that were specified to select patients that were included in the main database of the CDS-RA framework.

PAGNI Hospital Registry Database includes data from HeRBT (Hellenic Registry of Biologic Therapies), a nationwide multicenter prospective observational cohort of patients with inflammatory arthritis from seven healthcare centers in Greece (the PAGNI at Crete, the NIMITS at Athens, the Euroclinic at Athens, the Laikon at Athens, the Sismanoglio at Athens, the PGNI at Ioannina and the General Hospital Kavalas), established in 2004 [85]. RA, spondyloarthritis and other inflammatory arthritis patients treated with biologic agents in the participating centers are included in HeRBT at the initiation of their first biologic therapy (bDMARD) and are followed-up irrespectively of bDMARD treatment switches for as long as they receive bDMARDs. The decision to start bDMARD or to switch therapy is taken by the treating physician, based on national guidelines. Protocol approvals were obtained by local institutional review boards and all participants provided informed consent.

Clinical assessments are recorded every 3-6 months in the first 3 years and yearly afterwards while adverse events are reported whenever they occur. At baseline, recorded RA-related variables include demographics (age, gender) and disease characteristics (disease duration, previous anti-rheumatic drugs, rheumatoid factor, anti-CCP seropositivity, presence of erosions on X-Rays). At baseline and follow-up assessments, recorded variables include disease activity measures (DAS28, disease activity score 28; TJC28, tender joint count 28; SJC28, swollen joint count 28; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein; Physician VAS global, visual analogue score for global symptoms physician's assessment; Patient VAS global; visual analogue score for global symptoms patient's assessment; Patient VAS pain; visual analogue score for pain patient's assessment), function (HAQ, health assessment questionnaire), quality of life (Euroqol-5D), treatments (ongoing bDMARD, csDMARDs and prednisolone) and their dosage information. All adverse events are recorded by treating physicians prospectively at the time they occur and are classified according to seriousness based on the Common Terminology Criteria for Adverse Events (CTCAE). All treatment modifications are also tracked prospectively and if the bDMARD therapy is discontinued reasoning is provided.

The eligibility criteria specified to select patients included in the main database of the CDS-RA framework (the "RA Cohort on Biologics Database", Figure 3) from the PAGNI Hospital Registry Database were the following, patients should be adults diagnosed with RA and treated on bDMARDs that had been included until 31/5/2013 and had continuous follow-up for at least 5 years, irrespective of bDMARD switches. Data were censored when patients had completed 5 years of follow-up or at 30/6/2018, whichever occurred first. Patients having proportionally more missing than existing disease activity data in their 5-year follow-up have been excluded in order to maintain data quality, while missing disease activity data in the rest of the patients have been imputed.

2.4 Artificial Intelligence Methods for Medical Data Analysis

This section describes the statistical, machine learning and logical analyses of medical data that are performed in the Artificial Intelligence (AI) Layer of the CDS-RA framework architecture (Figure 3). Specifically, the AI layer includes (a) **Preparatory Analyses** to preprocess the patient and clinician data, (b) **Association Analyses of Difference** to compare patients groups of different persistent disease levels with regard to cumulative serious adverse events (SAEs) and functionality (HAQ), at 5 years of therapy follow-up, (c) **Explanatory**

Association Analyses to explain the long-term (5-year) trajectory of functionality based on patients' persistent disease level group and other clinical parameters, (d) **Predictive Analysis** to enable a personalized prediction of the persistent disease level group for a specific patient using his early therapy data (performance evaluation of the predictive analysis is also performed) and (f) **Logical Analysis** (Logical Reasoning) to derive a prognosis for a patient's persistent disease level based on many medical evidence sources and prioritization between different medical reasoning policies. All the analyses of the AI Layer are explained below in detail.

2.4.1 Preparatory Analyses

The Artificial Intelligence (AI) Layer of the CDS-RA system framework (Figure 3) supports preparatory analyses of the patients' and clinicians' data from the main database (the "RA Cohort on Biologics Database", Figure 3) in order to preprocess data for further analyses.

Specifically, the preparatory analyses focus on the following tasks, (a) the summarization of patients' longitudinal data (disease activity and functionality), in clinically significant follow-up time intervals, (b) the imputation of missing data in patients' follow-up time and (c) the categorization of patients based on their persistent disease activity level (PDL) into LDA (persistent low disease), MDA (persistent medium disease) and HDA (persistent high disease) groups, including the sub-categorization of MDA patients into lower-MDA (persistent lower medium disease) and higher-MDA (persistent higher medium disease) subgroups. Summary descriptive measures were used to represent the characteristics of patients in each group or subgroup, at therapy baseline and at specific therapy time intervals.

2.4.1.1 Longitudinal Data Summarization

In this study, the composite DAS28 score has been used to determine the following specific levels of disease activity, low/remission (DAS28<3.2), moderate (3.2≤DAS28≤5.1) and high (DAS28>5.1) [111]. Patients' longitudinal data have been summarized in therapy time (TT) intervals of clinical significance. Specifically, DAS28 and HAQ averages have been computed for every 6 months in the first 3 years of follow-up and yearly afterwards (TT intervals: months 3-8, 9-14, 15-20, 21-26, 27-32, 33-41, 42-53, 54-65, in total 8 TT intervals during 5 years of follow-up), providing thus a longitudinal representation of disease activity

and function trajectories over time, respectively. First semester's data analysis begins after the first therapy trimester which has been excluded as treatment adaptation period.

2.4.1.2 Missing Longitudinal Data Imputation

A minimal multivariate mixed-effect regression model has been applied for DAS28 imputation. DAS28 was modeled as the outcome dependent variable. Each patient had 9 DAS28 values representing DAS28 at baseline and the DAS28 averages in the 8 TT intervals. Time in treatment course was modeled as fixed effect variable (categorical dummy-coded of 9 values representing the baseline time point and the 8 TT intervals). The model accounted for individual patient variability (repeated patient measurements) with a random effect on the patient level (random intercept for each patient) and adjusted for possible confounding on the following fixed effect variables, gender (categorical variable with values female/male), age (quantitative variable with unit year) and disease duration (quantitative variable with unit year). The analysis was performed in MATLAB 9.2 statistical toolbox [107].

2.4.1.3 Persistent Disease Level (PDL) Patient Groups

Persistent disease level (PDL) was defined as having patient DAS28 in a specific range for at least 4 out of the 8 TT intervals (any of them) or equivalently for cumulative time percentage ≥50% of the 5-year follow-up time. *Cumulative time percentage* (CTP) of a DAS28 range was defined as the ratio of TT intervals that have DAS28 in the specific range, irrespective of fluctuations.

Patients were classified as *LDA*, (persistent low disease), *MDA* (persistent medium disease) and *HDA* (persistent high disease) persistent disease level group, if they correspondingly fulfilled the criterion of having, DAS28≤3.2, 3.2<DAS28≤5.1 and DAS28>5.1, for CTP≥50%. Patients in the MDA group were additionally classified as *lower-MDA* (persistent lower medium disease) and *higher-MDA* (persistent higher medium disease) subgroup, if they correspondingly fulfilled the additional criterion of having DAS28<4.2 and DAS28≥4.2, for CTP≥50%. Figure 4 depicts the hierarchy of persistent disease level groups.

The patients of the CDS-RA main database were categorized in the groups LDA, MDA and HDA based on their longitudinal prospectively recorded values of DAS28 (averaged in the TT intervals). Subsequently, the patients of the MDA group were sub-categorized into lower-MDA and higher-MDA subgroups. Conflicting cases where patients fulfilled the criteria of two different groups (half of the DAS28 averages within a group-

specific range and the others in another group's specified range) were resolved by the *worst case scenario policy*. This policy specified that in case of conflict, the group with the higher disease activity level should be preferred. This policy is selected in order to maximize the identification of patients with high disease levels. For example, patients that fulfilled the criteria of both HDA and any other group were always categorized into the HDA group while patients that fulfilled the criteria of both LDA and any other group, were always categorized in the latter.

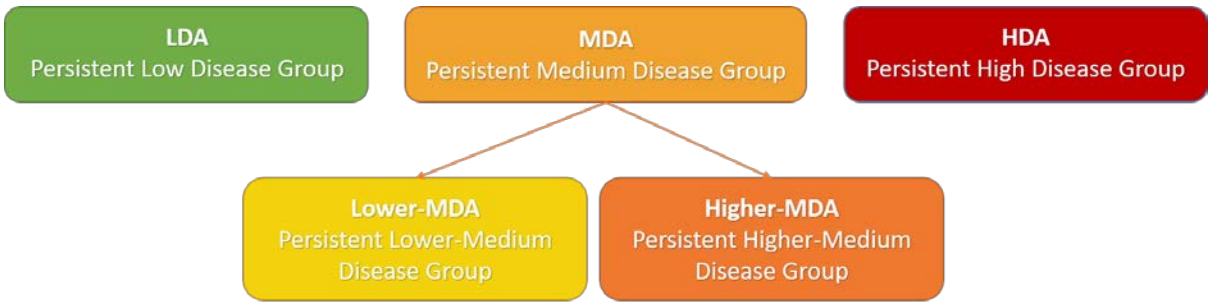


Figure 4. The Hierarchy of Persistent Disease Level Groups.

Summary descriptive measures were used to describe PDL groups baseline characteristics and the trajectories of disease activity, functionality and serious adverse events in the 5 years of follow-up time.

2.4.2 Association Analyses of Difference

The AI Layer of the CDS-RA system framework (Figure 3) supports association analyses of difference, in order to compare patient PDL groups with regard to functionality (HAQ) and serious adverse events (SAEs). Specifically, the association analyses of difference, focus on the comparison between the three PDL groups, LDA, MDA and HDA and between lower-MDA and higher-MDA subgroups, regarding the functionality status and cumulative SAEs that occurred at 5 years of therapy. In addition, comparisons are also performed between the groups on their baseline characteristics.

Non-parametric hypothesis tests, specifically Kruskal-Wallis test (for 3-group comparison of quantitative variable), Wilcoxon rank-sum test (for 2-group comparison of quantitative variable), and Chi-squared test (for 2-group or 3-group comparison of categorical variable), were used for the comparison of the groups. Groups were compared at baseline for

>50 and <100 characteristics, thus hypothesis tests used Bonferroni-corrected p-value=0.0019 to account for multiple comparisons.

2.4.3 Explanatory Association Analyses

The AI Layer of the CDS-RA system framework (Figure 3) supports explanatory association analyses in order to explain the longitudinal trajectory of patient functionality during 5 years of clinical follow-up based on their persistent disease level group and other clinical parameters.

Specifically, two explanatory association analysis were performed. The first explained patients' functionality during 5 years of clinical follow-up using information about their categorization into LDA, MDA, and HDA groups, their therapy time interval at the time of repeated monitoring and their age, gender and disease duration data. The second focused on MDA patients explaining their functionality during 5 years of clinical follow-up using information about the patients' categorization into lower-MDA and higher-MDA subgroups, their therapy time interval at the time of repeated monitoring and their age, gender and disease duration data.

Two multivariate linear mixed effect regression models were developed to perform the two explanatory association analysis. HAQ (patients repeated functionality measurements) was modeled as the dependent outcome variable in both models. The first model included patients categorized in LDA, MDA, and HDA groups and their group was modeled as fixed-effect variable (categorical variable, dummy-coded, reference-category LDA). The second model included only MDA patients that were sub-categorized into lower-MDA and higher-MDA subgroups and their subgroup was modeled as fixed-effect variable (categorical, dummy-coded, reference-category lower-MDA). Both models accounted for individual patient variability with a random effect on the patient level (random intercept for each patient) and adjusted for possible confounding effects of gender (binary variable with values female/male), age (quantitative variable with unit year) and disease duration (quantitative variable with unit year) that were modeled as fixed-effect variables. Time in treatment course was modeled in both models as fixed-effect categorical variable (dummy-coded, 9-values, reference-category baseline, remaining values represent the 8 TT intervals). This time representation was selected to model the change in HAQ in each TT interval compared to baseline. Alternative representations of groups' interactions with time did not yield significantly better results on comparison metrics of AIC, BIC and negative log-likelihood.

Essentially, the aim of these multivariable models is to analyze whether the modeled patient groups are associated with significantly different 5-year trajectory of patient functionality which will validate their heterogeneity and also the importance of patient PDL group prognosis.

2.4.4 Predictive Analysis

The AI Layer of the CDS-RA system framework (Figure 3) supports predictive analysis in order to enable a personalized prediction of the PDL group for a specific patient using their early therapy data from baseline and first therapy semester (first semester begins after the first therapy trimester which has been excluded as treatment adaptation period). Performance evaluation of the predictive analysis is also performed.

2.4.4.1 Early Prediction of Patient PDL Group

Two machine learning models were developed based on binary multivariable logistic regression in order to predict patient classification in the three PDL groups, LDA, MDA, and HDA. The first model focused on patient classification between LDA and the rest of the groups (MDA and HDA) in order to distinguish patients that are more likely to develop persistent low disease (LDA) than persistent medium (MDA) or high (HDA) disease. The second model is complementary to the first and focused on patient classification between MDA and HDA groups in order to distinguish patients that are more likely to develop persistent medium (MDA) than high (HDA) disease.

Both models adjusted for possible confounding effects of the following variables:

- *Gender*: Patient gender that was modeled as binary variable with values, 1=[Female] and 0=[Male].
- *Age*: Patient age that was modeled as quantitative variable with unit year.
- *Disease Duration <2 Years*: Patient disease duration that was modeled as binary variable with values, 1=[duration<2 years] and 0=[duration≥2 years]. Disease Duration was analyzed in this categorical form since this discretization was differentiated significantly between the three groups, LDA, MDA, and HDA based on Chi-squared hypothesis test while disease duration in the quantitive form (with unit year) was not differentiated between the three groups based on Kruskal-Wallis hypothesis test.

Models also included the following information about patient treatments at baseline and first therapy semester (first semester begins after the first therapy trimester which has been excluded as treatment adaptation period):

- *Previous csDMARDs Baseline Count <2*. Patient previous csDMARDs count at baseline that was modeled as binary variable with values, 1=[csDMARDs<2] and 0=[csDMARDs≥2] and was included in both models. Previous csDMARDs variable was analyzed in this categorical form since this discretization was univariably associated with group classification in both models compared to previous csDMARDs in the quantitive form that was not associated.
- *BTs 1st Semester Count >1*. Patient bDMARDs count in first semester that was modeled as binary variable with values, 1=[bDMARDs>1] and 0=[bDMARDs≤1] and was included in both models. Cumulative bDMARDs was analyzed in this categorical form since this discretization was univariably associated with group classification in both models compared to cumulative bDMARDs in the quantitive form that was not associated.
- *Anti-TNF Baseline.* Patient anti-TNF treatment initiation at baseline that was modeled as binary variable with values, 1=[anti-TNF initiated] and 0=[anti-TNF not initiated] and was included only in the first model since it was univariably associated with group classification only in the first model.
- *Prednisolone Baseline.* Prednisolone inclusion in patient treatment at baseline that was modeled as binary variable with values, 1=[prednisolone included] and 0=[prednisolone not included] and was included only in the second model since it was univariably associated with group classification only in the second model.

In addition, the models included the following information about patient disease activity, functionality and serious adverse events at baseline and first therapy semester (first semester begins after the first therapy trimester which has been excluded as treatment adaptation period):

- *DAS28 Baseline*. Patient disease activity measured with DAS28 score at baseline that was modeled as quantitative variable and was included in both models since it was univariably associated with group classification in both models.

- *HAQ Baseline.* Patient functionality measured with HAQ score at baseline that was modeled as quantitative variable and was included in both models since it was univariably associated with group classification in both models.
- *DAS28 1st Semester*. Patient average disease activity from DAS28 measurements in first semester that was modeled as quantitative variable and was included in both models since it was univariably associated with group classification in both models.
- *ΔHAQ 1st Semester <0.* Patient functionality improvement in first semester compared to baseline that was computed as the difference of the average patient HAQ score in first semester from patient baseline HAQ score and was modeled as binary variable with values, 1=[difference<0] and 0=[difference≥0]. Variable was included in both models since it was univariably associated with group classification in both models.
- *SAEs 1st Semester Count >0*. Patient serious adverse events in first semester that was modeled as binary variable with values, 1=[*SAEs*>0] and 0=[*SAEs*=0] and was included only in the second model since it was univariably associated with group classification only in the second model.

2.4.4.2 Predictive Performance Evaluation

Predictive efficiency of the two models was evaluated using the 10-fold cross-validation process (90% training set, 10% test set and 10 repetitions without resubstitution in test sets) in the samples of the analysis cohort, in order to avoid overfitting and selection bias. The following efficiency metrics were averaged from the 10 independent datasets extracted from the cross-validation process, Accuracy metric (ACC) specified as the rate of successful predictions), Sensitivity or True Positive Rate metric (TPR) specified as the rate of successful predictions for the first category in the binary classification model, Specificity or True Negative Rate (TNR) metric specified as the rate of successful predictions for the second category in the binary classification model and the Area Under the Receiver-Operating-Characteristic Curve metric (AUC).

2.4.5 Logical Analysis (Reasoning)

Clinicians decision-making process may include various clinical methodologies and medical evidence sources to derive conclusions on patient cases. Logical analysis (reasoning) is able to express formally with logical rules, the medical reasoning logic for clinical decision-

making in order to enable inference of meaningful clinical conclusions. In this study, it was developed a logical rule theory which expresses a medical reasoning logic to infer a prognosis for a patient's PDL group (LDA, MDA or HDA). The developed logical rule theory is loaded into an artificial intelligence (AI) engine integrated in the CDS-RA system (Figure 3, AI Layer of the CDS-RA system framework) to enable logical analysis (reasoning) and infer meaningful conclusions.

The developed logical rule theory is based on Defeasible Logic (Chapter I, section 1.4.3) and includes defeasible rules to support a prognosis of a patient's PDL group (LDA, MDA or HDA). The theory supports initially three medical reasoning policies for patient PDL group prognosis which are based on different medical evidence sources, namely *Prognostic Policy 1, 2* and *3.* The Prognostic Policies 1, 2 and 3 enable correspondingly, (a) inference of a patient's PDL group based on existing long-term follow-up data that fulfill a group's classification criteria by definition, (b) inference of a patient's PDL group based on clinicians' expert opinion and (c) inference of a patient's PDL group based on a predictive analysis from the patient's early (first treatment semester) therapy data.

The rule-based theory that is loaded into CDS-RA AI engine, is accessible, reusable, configurable and extendable with additional rules and policies to address the needs of specific medical domains and environments. The engine supports logical analysis (reasoning) with incomplete or contradictory medical evidence (inference of conclusions in the absence of information or in the presence of information conflicts, respectively). The developed rule-based theory includes conflict resolution rules to address patient cases which may be associated with contradictory conclusions on their PDL group prognosis, derived from the three Prognostic Policies. The three Prognostic Policies 1, 2 and 3 of the theory and their associated rules are described in the following sections in detail.

2.4.5.1 Prognostic Policy 1

The first policy namely *Prognostic Policy 1* (rules presented below), focuses on the inference of the patient's PDL group when the patient fulfills the criteria for group classification in LDA, MDA or HDA groups by their definition. Specifically, Prognostic Policy 1 includes three defeasible logical rules labeled "*rule_1*", "*rule_2*" and "*rule_3*". The first "*rule_1*" specifies the following knowledge, "A patient belongs to LDA group when the patient has an average DAS28 score within remission or low range for 4 TT intervals (any of them)". The second "*rule_2*" expresses the following knowledge, "A patient belongs to MDA group when

the patient has an average DAS28 score within moderate range for 4 TT intervals (any of them).". Finally, the third "*rule_3*" defines the following knowledge, "A patient belongs to HDA group when the patient has an average DAS28 score within high range for 4 TT intervals (any of them).". The patient is specified by unique identifier variable ID.

$\boldsymbol{rule\_1}$: $patient\_has\_4\_TT\_average\_DAS\_in\_range(ID, remission\_or\_low)$
$\Longrightarrow belongs(ID, groupLDA)$

$\boldsymbol{rule\_2}$: $patient\_has\_4\_TT\_average\_DAS\_in\_range(ID, moderate)$
$\Longrightarrow belongs(ID, groupMDA)$

$\boldsymbol{rule\_3}$: $patient\_has\_4\_TT\_average\_DAS\_in\_range\ (ID, high)$
$\Longrightarrow belongs(ID, groupHDA)$

$\boldsymbol{rule\_4}$: $\rightarrow conflicting\_pairs ::$
$belongs(ID, groupHDA), belongs(ID, groupMDA), belongs(ID, groupLDA)$

$\boldsymbol{rule\_5}$: $\rightarrow superior(rule\_3, rule\_1)$

$\boldsymbol{rule\_6}$: $\rightarrow superior(rule\_3, rule\_2)$

$\boldsymbol{rule\_7}$: $\rightarrow superior(rule\_2, rule\_1)$

Patients follow-up time spans 8 TT intervals. Thus, there might exist patient cases that fulfill the preconditions of more than one rule of the aforementioned rules (having 4 TT intervals in one range and the rest in another) leading to the conclusion that they belong to more than one PDL group. In such cases, it is evident that a conflict resolution policy is required in order to infer a unique PDL group. Prognostic Policy 1 includes rule labeled "*rule_4*" (Conflicting Literals of Defeasible Logic, Chapter I, section 1.4.3) to specify the conflict that arises when more than one PDL group is derived for the same patient. Specifically, the rule expresses the following knowledge "If there is evidence that a patient with unique identifier ID belongs to two PDL groups from groups LDA, MDA, and HDA then this patient case will lead to conflicting information about his PDL group.".

In the developed theory, a conflict will arise if the preconditions of two out of the three rules "*rule_1*", "*rule_2*" and "*rule_3*" are met for a specific patient since these rules' conclusions can be associated with the conflicting information defined on "*rule_4*". Therefore, Prognostic Policy 1 includes additional Priority Rules labeled "*rule_5*", "*rule_6*"

and "*rule_7*" that define a conflict resolution policy (Priority rules of Defeasible Logic, Chapter I, section 1.4.3) between rules *"rule_1"*, *"rule_2"* and *"rule_3"* using priorities. The Priority Rules define that the conclusion of *"rule_3"* should be preferred compared to the rules *"rule_1"* and *"rule_2"* while the conclusion of *"rule_2"* should be preferred compared to *"rule_1"*. Essentially, these Priority Rules support the following conflict resolution policy of the worst case scenario, "If a patient is associated with evidence for both HDA group and another group (MDA or LDA) based on his average DAS28 score of 8 TT intervals, then HDA should be preferred. Additionally, if a patient is associated with evidence for both MDA and LDA group based on his average DAS28 score of 8 TT intervals, then MDA should be preferred.".

2.4.5.2 Prognostic Policy 2

The second policy, namely *Prognostic Policy 2* (rules presented below), focuses on the inference of the patient's PDL group based on clinicians' expert opinion. Specifically, Prognostic Policy 2 includes the defeasible rule labeled *"rule_8"* that defines the following knowledge, "If a clinician with unique identifier Clinician_ID classifies a patient with unique identifier Patient_ID into a PDL group (LDA, MDA or HDA) then it is inferred that the patient belongs to this group". In addition, in this theory we want to express that *Prognostic Policy 1* provides stronger evidence than *Prognostic Policy 2* on a patient's PDL group. This prioritization between the Prognostic Policies 1 and 2 is expressed by priority rules, *"rule_9"*, *"rule_10"* and *"rule_11"*. Essentially, the priorities of the aforementioned rules specify that the conclusions of *"rule_1"*, *"rule_2"* and *"rule_3"* from *Prognostic Policy 1* provide stronger evidence than the conclusion of *"rule_8"* from *Prognostic Policy 2* in case of a conflict for a specific patient's PDL group (as defined in "*rule_4*" Conflicting Literals rule).

$\boldsymbol{rule\_8}$: $expert\_clinician\_classification\,(Clinician\_ID, Patient\_ID, Group)$
$\Longrightarrow\ belongs(Patient\_ID, Group)$

$\boldsymbol{rule\_9}$: $\rightarrow\ superior(rule\_3, rule\_8)$
$\boldsymbol{rule\_10}$: $\rightarrow\ superior(rule\_2, rule\_8)$
$\boldsymbol{rule\_11}$: $\rightarrow\ superior(rule\_1, rule\_8)$

2.4.5.3 Prognostic Policy 3

The third policy namely *Prognostic Policy 3* (rules presented below), focuses on the inference of the patient's PDL group based on a predictive analysis from the patient's early (first treatment semester) therapy data. The predictive analysis includes two binary classification models, the first focused on patient classification between LDA and the rest of the groups (MDA and HDA) while the second is complementary to the first and focused on further classification between MDA and HDA groups. Prognostic Policy 3 includes the defeasible rule labeled "rule_12" that defines the following knowledge, "If the first predictive model of patient classification between LDA and the rest of the groups (MDA and HDA), classified a patient with unique identifier Patient_ID into the LDA group, then it is inferred that the patient belongs to LDA group.". In addition, Prognostic Policy 3 includes the defeasible rule labeled "rule_13" that defines the following knowledge, "If the first predictive model (LDA vs the rest of the groups MDA and HDA), did not classify a patient with unique identifier Patient_ID into the LDA group (note the negation symbol ¬ on the precondition "$\neg predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA)$") and also the second predictive model classified the patient into a group between MDA and HDA, then it is inferred that the patient belongs to this group.".

$\boldsymbol{rule\_12}$: $predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA)$
$\Longrightarrow belongs(Patient\_ID, groupLDA)$

$\boldsymbol{rule\_13}$: $\neg\, predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA),$
$predict\_MDA\_vs\_HDA(Patient\_ID, Group)$
$\Longrightarrow belongs(Patient\_ID, Group)$

$\boldsymbol{rule\_14}$: $\rightarrow superior(rule\_1, rule\_12)$
$\boldsymbol{rule\_15}$: $\rightarrow superior(rule\_2, rule\_12)$
$\boldsymbol{rule\_16}$: $\rightarrow superior(rule\_3, rule\_12)$
$\boldsymbol{rule\_17}$: $\rightarrow superior(rule\_1, rule\_13)$
$\boldsymbol{rule\_18}$: $\rightarrow superior(rule\_2, rule\_13)$
$\boldsymbol{rule\_19}$: $\rightarrow superior(rule\_3, rule\_13)$
$\boldsymbol{rule\_20}$: $\rightarrow superior(rule\_8, rule\_12)$
$\boldsymbol{rule\_21}$: $\rightarrow superior(rule\_8, rule\_13)$

In addition, in this theory we want to express that *Prognostic Policy 1 and 2* provide stronger evidence than *Prognostic Policy 3* on a patient's PDL group since Prognostic Policy

3 is based only on patients' early therapy data (first treatment semester) while the other policies (Prognostic Policies 1 and 2) utilize patients' long-term therapy data and clinician insights, respectively. Prognostic policy 3 is thus more appropriate for patient cases without long-term follow-up or clinical insights. This prioritization between the Prognostic Policies 1, 2 and 3 is expressed by priority rules 14-21 (labels *"rule_14"*, *"rule_15"*, …, *"rule_21").* Essentially, the priorities of the rules 14-19 specify that the conclusions of *"rule_1"*, *"rule_2"* and *"rule_3"* from *Prognostic Policy 1* provide stronger evidence than the conclusions of *"rule_12"* and *"rule_13"* from *Prognostic Policy 3*, in case of a conflict for a specific patient's PDL group (as defined in "*rule_4*" Conflicting Literals rule). Furthermore, the priorities of the rules 20-21 specify that the conclusion of *"rule_8"* from *Prognostic Policy 2* provides stronger evidence than the conclusions of *"rule_12"* and *"rule_13"* of *Prognostic Policy 3*, in case of a conflict.

3. Results

The development results of the CDS-RA system framework are provided in the following sections. Section 3.1 describes the characteristics of the entire patient cohort that was included in the data analyses of the system. Section 3.2 describes the characteristics of the three distinct patient groups with persistent low (LDA), medium (MDA) and high (HDA) disease and the two MDA patient subgroups (lower-MDA and higher-MDA), regarding their baseline status, their disease activity courses and the cumulative biologic treatments during therapy. Section 3.3 provides the results from the association analyses of difference supported by the AI Layer of the CDS-RA system which analyze the difference in serious adverse events occurrence between patient groups LDA, MDA and HDA and between patient subgroups lower-MDA and higher-MDA. Section 3.4 describes the results from the explanatory association analyses supported by the AI Layer of the CDS-RA system that model the longitudinal 5-year trajectory of patients' functionality (HAQ) based on their persistent disease level (PDL) group and other clinical parameters. The analyses explain the different effect of the PDL groups (LDA, MDA, and HDA) and subgroups (lower-MDA and higher-MDA) in the patients' 5-year functionality course. Section 3.5 presents the results from the predictive analyses supported by the AI Layer of the CDS-RA system which provide early predictors for patient classification in the persistent disease level groups (LDA, MDA, and HDA) using early patient therapy data. Section 3.6 describes the results from the development of the AI rule theory that supports logical analysis (reasoning) on the patient

group prognosis (LDA, MDA, and HDA). Finally, section 3.7 illustrates the entire graphical user interface of the CDS-RA system, emphasizing on the specific AI components which perform the prognosis of a patient's PDL group (LDA, MDA, or HDA) based on logical analysis of formal rule-based medical reasoning policies (described in the methodology section 2.4.6) that include different prioritized sources of medical evidence.

3.1 Patient Cohort Characteristics

The analysis cohort included 527 RA patients out of the 1466 recorded in the registry until May 2013. This selected cohort was the result of a strict screening process focused on data quality that sequentially excluded 763 patients having <5 years of follow-up, 166 patients having >50% missing longitudinal DAS28-data and 10 patients not exhibiting any persistent disease activity level. The cohort had a mean monitoring duration of 8 ±2.6 years. Baseline characteristics of the cohort are described in Table 4. This was a typical cohort of RA patients treated on bDMARDs with established disease (mean disease duration 9.31 ±8.59 years), 16% early arthritis (disease duration <2 years), high disease activity (mean DAS28 6 ±1.1), and moderate functional limitation (mean HAQ 0.95 ±0.5). Patients were treated with a mean of 2.39 (±1.1) csDMARDs prior to the first bDMARD, while 71% were on combination with methotrexate and 43% on prednisolone drug. At the end of the 5-year follow-up, a total of 334 (63%) patients received a second sequential bDMARD, while 203 (37%) received a third bDMARD (detailed information on specific biologic agents is provided in Appendix B. Supplementary Table 1).

Table 4. Baseline Characteristics of Cohort and Patient Groups LDA, MDA, and HDA.

Variable Name	**Cohort (n=527)**	**LDA (n=90)**	**MDA (n=295)**	**HDA (n=142)**	**p-value ‡**
Females	418 (79%)	52 (58%)	241 (82%)	125 (88%)	p<0.0001*[(a)]
Age (years)	57.02 ±12	51.96 ±13	57.69 ±12	58.83 ±10	p=0.0009*[(a)]
RA Duration (years)	9.31 ±8.59	9.57 ±10	9.1 ±8.12	9.56 ±8.56	p=0.45[(b)]
RA Duration <2 years	82(16%)	22(24%)	48 (16%)	12 (8%)	p=0.0041
Seropositive (n=343) **	150 (44%)	15 (47%)	85 (46%)	59 (40%)	p=0.51
TJC28 (n=440) **	12.5 ±7.45	7.93 ±6.3	11.14 ±6.4	17.53 ±7.4	p<0.0001*[(c)]
SJC28 (n=440) **	10.64 ±7.1	6.17 ±6.1	9.66 ±6.1	14.82 ±7.4	p<0.0001*[(c)]
ESR mm/s (n=438) **	38.43 ±24	35.37 ±23	39.4 ±24	37.88 ±24	p=0.41

CRP mg/dl (n=378) **	2.23 ±5.32	4.07 ±12	2.22 ±3.15	1.28 ±2.3	p<0.0001*[(c)]
DAS28 (Imputed)	6 ±1.1	4.92 ±1	5.94 ±0.86	6.8 ±0.97	p<0.0001*[(c)]
CDAI (n=403) **	36.42 ±15	26.06 ±12	34.6 ±12	47.38 ±14	p<0.0001*[(c)]
SDAI (n=346) **	38.42 ±15	29.99 ±17	36.64 ±12	47.86 ±15	p<0.0001*[(c)]
Physician VAS G. (n=407) **	70.73 ±15	62.92 ±16	71.13 ±14	74.08 ±14	p<0.0001*[(a)]
Patient VAS G. (n=439) **	67.86 ±19	59.43 ±25	67.69 ±17	72.28 ±19	p=0.0015[(b)]
Patient VAS Pain (n=419) **	69.14 ±19	62.93 ±25	68.5 ±17.2	73.62 ±18	p=0.0040[(b)]
HAQ (Imputed)	0.95 ±0.5	0.63 ±0.4	0.9 ±0.5	1.24 ±0.5	p<0.0001*[(c)]
Euroqol (n=127) **	0.26 ±0.4	0.36 ±0.4	0.34 ±0.39	0.11 ±0.39	p=0.01[(b)]
Previous csDMARDs	2.39 ±1.1	2.11 ±0.9	2.41 ±1.18	2.53 ±1.01	p=0.008[(a)]
Ongoing csDMARDs	1.18 ±0.6	1.15 ±0.6	1.15 ±0.6	1.25 ±0.6	p=0.32
Monotherapy bDMARD	38 (7%)	7 (8%)	25 (9%)	6 (4%)	p=0.25
Methotrexate	372 (71%)	62 (69%)	217 (74%)	93 (66%)	p=0.21
Anti-TNF	467 (89%)	88 (98%)	262 (89%)	117 (82%)	p=0.0015[(a)]
Prednisolone	224 (43%)	44 (49%)	140 (48%)	40 (28%)	p=0.0003[(b)]

Results presented as counts n (%) or mean (±sd).

** Missing data >5%.

‡ p-value of 3-group comparison test, Kruskal-Wallis or Chi-square test, as appropriate.

* Test of 3-group comparison between LDA, MDA, and HDA groups (Kruskal-Wallis or Chi-square test, as appropriate), yielded differentiation between the groups (p<0.05 threshold and p<0.0019 Bonferroni-corrected threshold).

[(a)] Wilcoxon rank-sum tests yielded that LDA group was differentiated from MDA and HDA groups (p<0.05 threshold).

[(b)] Wilcoxon rank-sum tests yielded that HDA group was differentiated from LDA and MDA groups (p<0.05 threshold).

[(c)] Wilcoxon rank-sum tests yielded differentiation in all pair-wise comparisons of LDA, MDA, and HDA groups (p<0.05 threshold).

Terms: TJC28=Tender Joint Count 28; SJC28= Swollen Joint Count 28; ESR=erythrocyte sedimentation rate; CRP=C reactive protein; DAS28=Disease Activity Score 28 with ESR4; CDAI=Clinical Disease Activity Index; SDAI=Simplified Disease Activity Index; VAS= Visual Analogue Scale 0-100; Physician VAS G.=Physician VAS Global; Patient VAS G.=Patient VAS Global; HAQ=Health Assessment Questionnaire; csDMARDs=Conventional Synthetic Disease-modifying Antirheumatic Drugs; Anti-TNF= Tumor Necrosis Factor Alpha (TNFα) Inhibitors; LDA= Persistent Low Disease; MDA=Persistent Medium Disease; HDA=Persistent High Disease.

3.2 Patient Groups Characteristics

The categorization of patients from the CDS-RA main database yielded 90 (17%) in LDA group, 295 (56%) in MDA group and 142 (27%) in HDA group. In addition, patients with MDA have been subdivided into 133 (25%) patients in lower-MDA subgroup and 162 (31%)

patients in higher-MDA subgroup. Importantly, results show that patients with persistent medium disease (MDA) comprise the majority of RA patients under biologic therapy in clinical practice. Section 3.2.1 describes the groups' (LDA, MDA, and HDA) and subgroups' (lower-MDA and higher-MDA) baseline characteristics, section 3.2.2 presents characteristics from their 5-year disease activity course and section 3.2.3 provides information on biologic treatments during 5-year therapy.

3.2.1 Patient Groups Baseline Characteristics

Baseline characteristics of the groups LDA, MDA, and HDA are provided in Table 4. Association analyses of difference yielded that patients in MDA group were older than in LDA group at baseline (Age years: 58 ±12 in MDA vs 52 ±13 in LDA; $p<0.05$) and more frequently females (Gender female: 82% in MDA vs. 58% in LDA; $p<0.05$), had worse physician global assessment (Physician VAS G: 71.13 ±14 in MDA vs 62.92 ±16 in LDA; $p<0.05$) and were treated with higher number of csDMARDs prior to the first bDMARD (csDMARDs: 2.41 ±1.18 in MDA vs 2.11 ±0.9 in LDA, $p<0.05$). In addition, patients in HDA group had worse patient global self-assessment at baseline (Patient VAS G.: 72.28 ±19 in HDA vs 67.69 ±17 in MDA, $p<0.05$) and worse patient pain self-assessment (Patient VAS Pain: 73.62 ±18 in HDA vs 68.5 ±17.2, $p<0.05$).

In general, higher PDL patient groups had worse disease activity at baseline (DAS28: 4.9 ±1.0 in LDA vs 5.9 ±0.9 in MDA vs 6.8 ±1.0 in HDA; $p<0.05$, similar results for CDAI and SDAI) and worse functionality status (HAQ: 0.63 ±0.4 in LDA vs 0.9 ±0.5 in MDA vs 1.24 ±0.5 in HDA; $p<0.05$), higher number of tender joints (TJC28: 7.93 ±6.3 in LDA vs 11.14 ±6.4 in MDA vs 17.53 ±7.4 in HDA, $p<0.05$), higher number of swollen joints (SJC28: 6.17 ±6.1 in LDA vs 9.66 ±6.1 in MDA vs 14.82 ±7.4 in HDA, $p<0.05$), lower levels of C-reactive protein (CRP mg/dl: 4.07 ±12 in LDA vs 2.22 ±3.15 in MDA vs 1.28 ±2.3 in HDA, $p<0.05$) and included less frequently early arthritis patients (Disease duration<2 years: 24% in LDA vs 16% in MDA vs 8% in HDA; $p<0.05$).

Baseline characteristics of the subgroups lower-MDA and higher-MDA are provided in Table 5. Association analysis of difference yielded that patients in higher-MDA subgroup were older than patients in lower-MDA subgroup at baseline (Age years: 60 ±12 in higher-MDA vs 55 ±12 in lower-MDA; $p<0.05$) and more frequently females (Gender female: 86% in higher-MDA vs 78% in lower-MDA; $p<0.05$), had higher number of tender joints (TJC28: 12.39 ±6.74 in higher-MDA vs 9.61 ±5.56 in lower-MDA; $p<0.05$) and higher number of

swollen joints (SJC28: 10.89 ±6.26 in higher-MDA vs 8.14 ±5.55 in lower-MDA; p<0.05). In addition, patients in higher-MDA subgroup had worse disease activity than patients in lower-MDA subgroup at baseline (DAS28: 6.02 ±0.86 in higher-MDA vs 5.62 ±0.74 in lower-MDA; p<0.05, similar results for CDAI and SDAI) and worse functionality status (HAQ: 1 ±0.45 in higher-MDA vs 0.77 ±0.43 in lower-MDA; p<0.05) and they also had worse physician global assessment (Physician VAS G: 73.97 ±12.7 in higher-MDA vs 67.57 ±15.12 in lower-MDA; p<0.05) and worse patient global self-assessment (Patient VAS G.: 70.93 ±15.54 in higher-MDA vs 63.74 ±17.82 in lower-MDA; p<0.05).

Table 5. Baseline Characteristics of Patient Subgroups lower-MDA and higher-MDA.

Variable Name	**Cohort (n=527)**	**MDA (n=295)**	**lower-MDA (n=133)**	**higher-MDA (n=162)**
Females [(a)]	418 (79%)	241 (82%)	102 (78%)	139 (86%)
Age (years) [(a)]	57.02 ±12	57.69 ±12	55.11 ±12.31	59.81 ±11.59
RA Duration (years)	9.31 ±8.59	9.1 ±8.12	8.71 ±7.71	9.43 ±8.45
RA Duration <2 years	82(16%)	48 (16%)	20 (15%)	28 (17%)
Seropositive (n=343) **	150 (44%)	85 (46%)	37 (51%)	48 (43%)
TJC28 (n=440) ** [(a)]	12.5 ±7.45	11.14 ±6.4	9.61 ±5.56	12.39 ±6.74
SJC28 (n=440) ** [(a)]	10.64 ±7.1	9.66 ±6.1	8.14 ±5.55	10.89 ±6.26
ESR mm/s (n=438) **	38.43 ±24	39.4 ±24	37.86 ±23.42	40.63 ±25.02
CRP mg/dl (n=378) **	2.23 ±5.32	2.22 ±3.15	2.31 ±3.32	2.15 ±3.02
DAS28 (Imputed) [(a)]	6 ±1.1	5.94 ±0.86	5.62 ±0.74	6.02 ±0.86
CDAI (n=403) ** [(a)]	36.42 ±15	34.6 ±12	30.86 ±10.05	37.62 ±12.13
SDAI (n=346) ** [(a)]	38.42 ±15	36.64 ±12	32.66 ±10.44	39.63 ±12.18
Physician VAS G. (n=407) ** [(a)]	70.73 ±15	71.13 ±14	67.57 ±15.12	73.97 ±12.7
Patient VAS G. (n=439) ** [(a)]	67.86 ±19	67.69 ±17	63.74 ±17.82	70.93 ±15.54
Patient VAS Pain (n=419) **	69.14 ±19	68.5 ±17.2	66.26 ±18.57	70.38 ±15.87
HAQ (Imputed) [(a)]	0.95 ±0.5	0.9 ±0.5	0.77 ±0.43	1 ±0.45
Euroqol (n=127) **	0.26 ±0.4	0.34 ±0.39	0.4 ±0.34	0.26 ±0.43
Previous csDMARDs	2.39 ±1.1	2.41 ±1.18	2.34 ±1.25	2.46 ±1.11
Ongoing csDMARDs	1.18 ±0.6	1.15 ±0.6	1.1 ±0.63	1.19 ±0.53
Monotherapy [(a)]	38 (7%)	25 (9%)	16 (12%)	9 (6%)

Methotrexate	372 (71%)	217 (74%)	97 (73%)	120 (74%)
Anti-TNF [(a)]	467 (89%)	262 (89%)	124 (93%)	138 (85%)
Prednisolone	224 (43%)	140 (48%)	71 (53%)	69 (43%)

Results presented as counts n (%) or mean (±sd).

** Missing data >5%.

[(a)] Wilcoxon rank-sum tests yielded differentiation between lower-MDA and higher-MDA ($p<0.05$ threshold).

Terms: TJC28=Tender Joint Count 28; SJC28=Swollen Joint Count 28; ESR=erythrocyte sedimentation rate; CRP=C reactive protein; DAS28=Disease Activity Score 28 ESR4; CDAI=Clinical Disease Activity Index; SDAI= Simplified Disease Activity Index; VAS=Visual Analogue Scale 100; Physician VAS G.=Physician VAS Global; Patient VAS G.=Patient VAS Global; HAQ=Health Assessment Questionnaire; csDMARDs=Conventional Synthetic Disease-modifying Antirheumatic Drugs; Anti-TNF=TNFα inhibitors. LDA= Persistent Low Disease; MDA= Persistent Medium Disease; HDA= Persistent High Disease.

3.2.2 Patient Groups Disease Activity Characteristics

A representation of the DAS28 course for each patient in the groups LDA, MDA, and HDA is provided in Figure 5. The LDA, MDA, and HDA patient groups' 5-year disease activity trajectories (average DAS28 in the 8 TT intervals) are presented in Figure 6 A., showing a clear distinct trajectory for each group. Figure 6 B. presents the lower-MDA and higher-MDA subgroups' 5-year disease activity trajectories (average DAS28 in the 8 TT intervals), showing also a clear distinct trajectory for each subgroup. Association analysis of difference yielded that higher PDL patient groups continued having worse disease activity at 5 years as compared to baseline (DAS28: 2.45 ±0.91 in LDA, 3.85 ±1 in MDA and 5.24 ±1.025 in HDA; $p<0.001$) which also applied between the two MDA subgroups, lower-MDA and higher-MDA (DAS28: 3.49±0.91 in lower-MDA and 4.14±0.95 in higher-MDA; $p<0.001$).

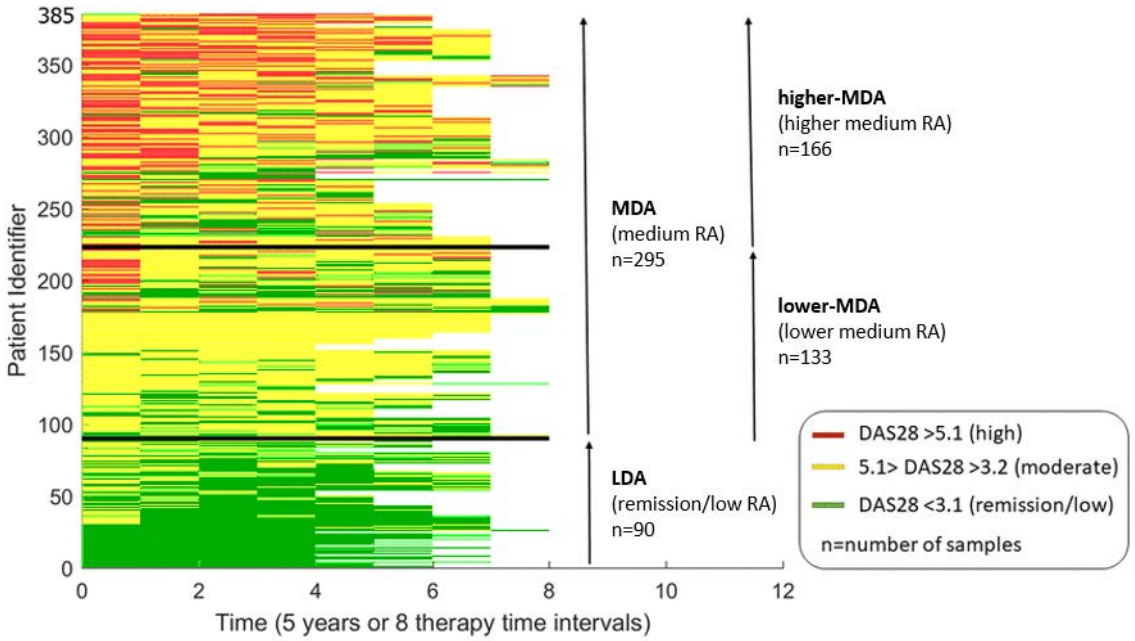


Figure 5. DAS28 5-Year Course for Each Patient of Groups (Imputed Data).

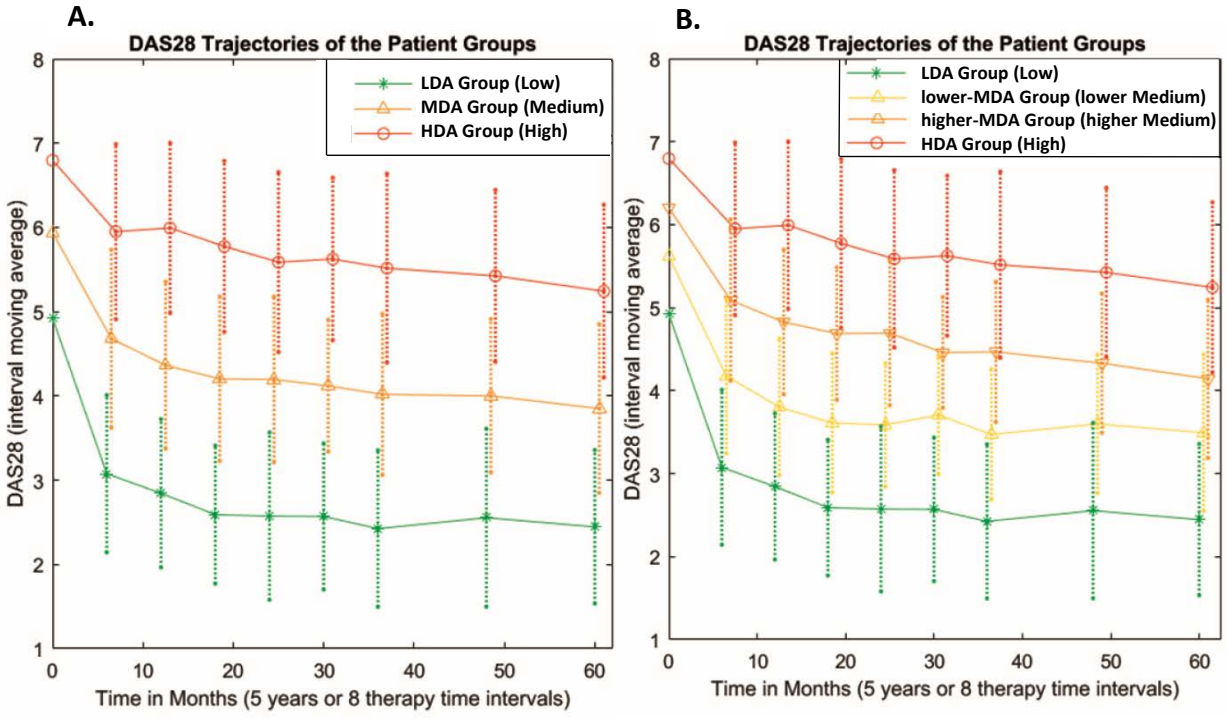


Figure 6. DAS28 Trajectories of the Patient Groups.

A. Persistent Low (LDA), Medium (MDA) and High (HDA) Disease.

B. Persistent Low (LDA), Lower Medium (lower-MDA), Higher Medium (higher-MDA) and High (HDA) disease.

3.2.3 Patient Groups Biologic Treatments Characteristics

The LDA, MDA, and HDA patient groups' 5-year cumulative biologic treatments (BTs) trajectories (treatments count from baseline to the end of each of the 8 TT intervals) are presented in Figure 7 A., showing a clear distinct trajectory for each group. Figure 7 B. presents the lower-MDA and higher-MDA subgroups' 5-year cumulative biologic treatments trajectories, showing also a clear distinct trajectory for each subgroup. Association analysis of difference yielded that at 5 years higher PDL patient groups had received a larger number of biologic treatments (BTs: 1.26±0.53 in LDA, 1.82 ±1.03 in MDA and 2.91 ±1.29 in HDA; $p<0.001$) which also applied between the two MDA subgroups, lower-MDA and higher-MDA (BTs: 1.49 ±0.82 in lower-MDA and 2.1 ±1.1 in higher-MDA; $p<0.001$).

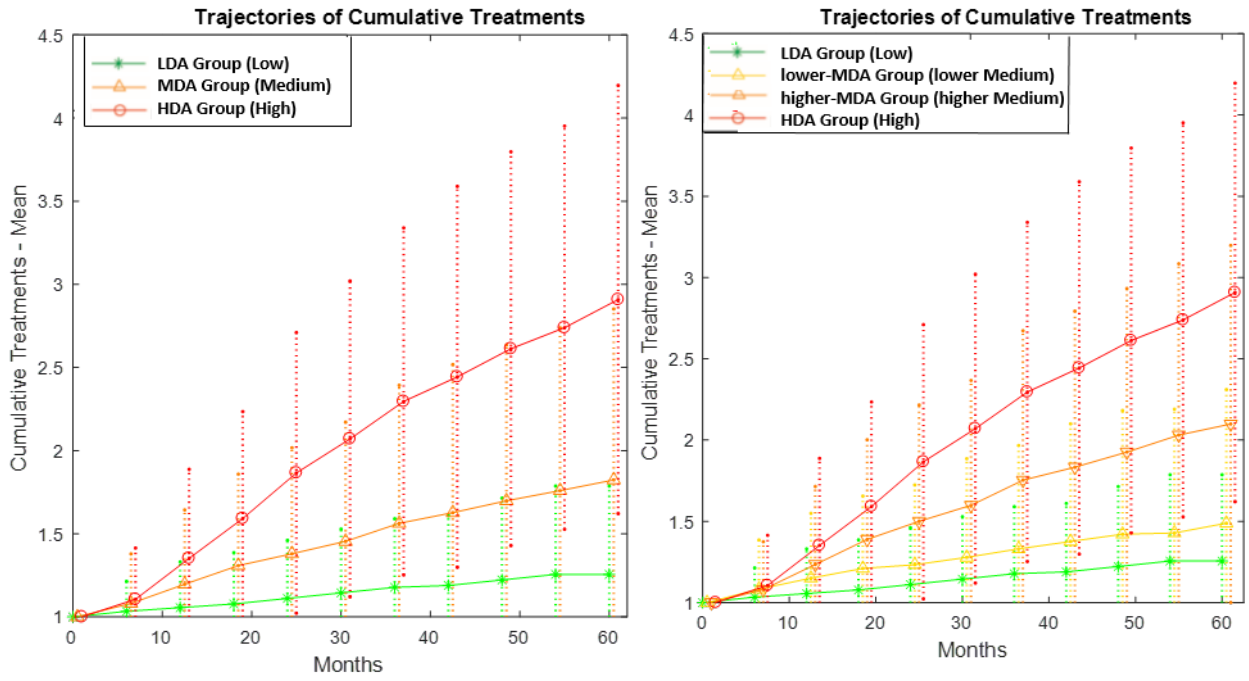


Figure 7. Cumulative Biologic Treatments Trajectories of the Patient Groups.

A. Persistent Low (LDA), Medium (MDA) and High (HDA) disease.

B. Persistent Low (LDA), Lower Medium (lower-MDA), Higher Medium (higher-MDA) and High (HDA) disease.

3.3 Patient Groups Serious Adverse Events Outcome

The LDA, MDA, and HDA patient groups' 5-year cumulative serious adverse events (SAEs) trajectories (SAEs count from baseline to the end of each of the 8 TT intervals) are presented in Figure 8 A. showing a clear distinct trajectory for each group. Figure 8 B. presents the lower-MDA and higher-MDA subgroups' 5-year cumulative SAEs trajectories showing also a clear distinct trajectory for each subgroup. Association analysis of difference yielded that at 5 years higher PDL patient groups exhibited a higher number of serious adverse events (SAEs: 0.2±0.48 in LDA, 0.5±0.96 in MDA and 0.89±1.7 in HDA; $p<0.01$) which also applied between MDA subgroups lower-MDA and higher-MDA (SAEs: 0.32 ±0.6 in lower-MDA and 0.64 ±1.16 in higher-MDA; $p=0.038$).

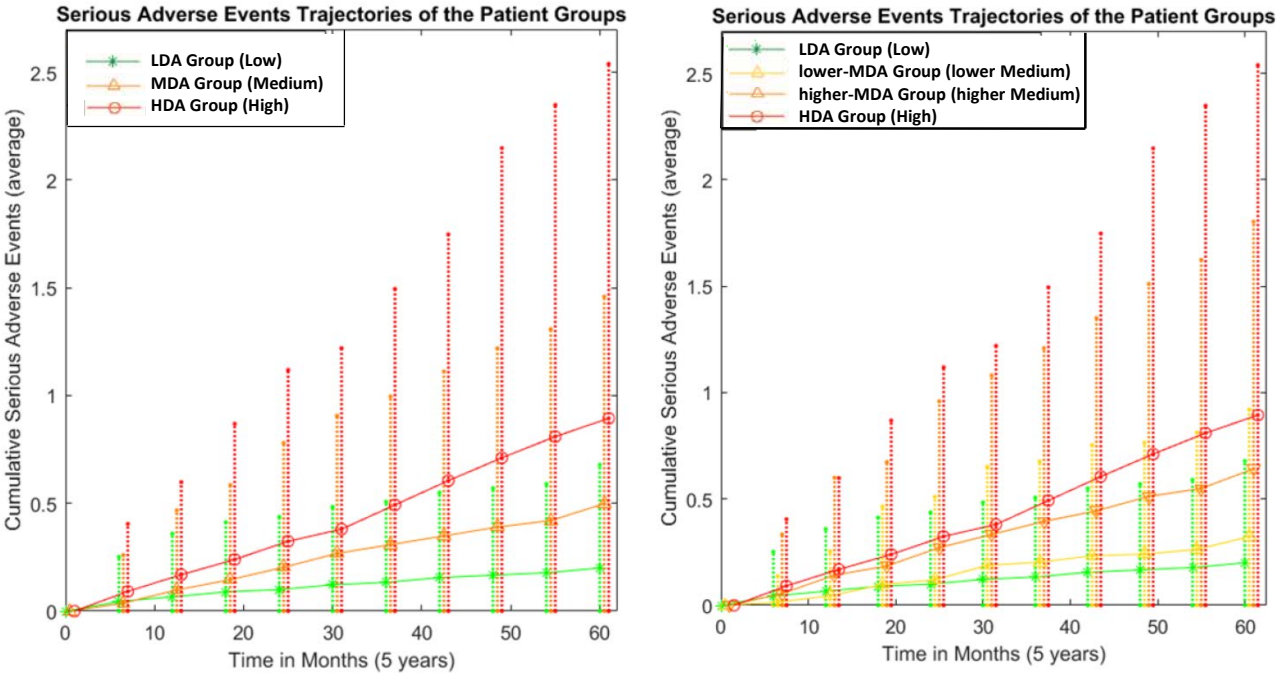


Figure 8. SAEs Trajectories of the Patient Groups.

A) Persistent Low (LDA), Medium (MDA) and High (HDA) Disease.

B) Persistent low (LDA), Lower Medium (lower-MDA), Higher Medium (higher-MDA) and High (HDA) Disease.

3.4 Patient Groups Functionality Outcome

The LDA, MDA, and HDA patient groups' 5-year functionality (HAQ) trajectories (average DAS28 in the 8 TT intervals) are presented in Figure 9 A. showing a clear distinct trajectory

for each group. Figure 9 B. presents the lower-MDA and higher-MDA subgroups' 5-year HAQ trajectories showing also a clear distinct trajectory for each subgroup. Association analysis of difference yielded that at 5 years higher PDL patient groups exhibited worse functionality (HAQ: 0.3±0.31 in LDA, 0.55±0.47 in MDA and 1±0.52 in HDA; $p<0.001$) which also applied for the two MDA subgroups, lower-MDA and higher-MDA (HAQ: 0.41 ±0.38 in lower-MDA and 0.66±0.52 in higher-MDA; $p<0.001$).

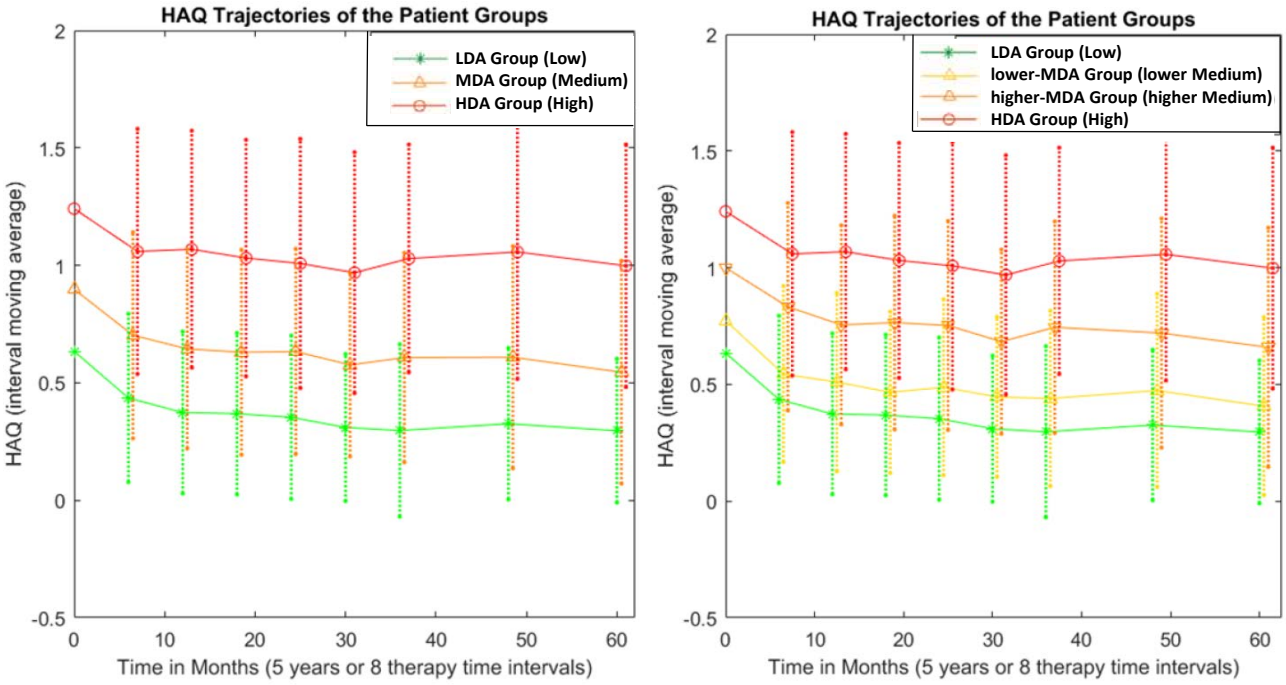


Figure 9. HAQ Trajectories of the Patient Groups.

A) Persistent Low (LDA), Medium (MDA) and High (HDA) Disease.

B) Persistent low (LDA), Lower Medium (lower-MDA), Higher Medium (higher-MDA) and High (HDA) Disease.

Multivariate linear mixed effect regression analysis (explanatory association analysis) was performed on patient groups LDA, MDA, and HDA to analyze variables associated with their longitudinal 5-year course of functionality. In multivariable mixed-effect regression analysis (Table 6), the MDA group was associated with worse 5-year functionality trajectory than the LDA group (+0.27 higher HAQ trajectory in MDA, 95% CI +0.22 to +0.33, $p<0.0001$) while the HDA group was associated with an even higher 5-year functionality limitation than the LDA group (+0.69 higher HAQ trajectory in MDA, 95% CI +0.63 to +0.75, $p<0.0001$). Analysis was adjusted for possible confounding effects on gender, age and

disease duration. The 5-year functionality was also significantly worse in females than males (+0.15 higher HAQ trajectory in females, 95% CI +0.1 to +0.19, p=0.0014) and in patients of older age (+0.01 higher HAQ trajectory per year, 95% CI +0.008 to +0.011, p<0.0001). At 5 years, all groups were associated with improved functionality as compared to baseline (HAQ decrease from baseline: -0.318 HAQ, 95% CI -0.287 to -0.349, p<0.0001). Notably, the improvement occurred mainly within the first 12 months of treatment (HAQ decrease from baseline in first 12 months: -0.234, 95% CI -0.202 to -0.267, p<0.0001), showing minor further improvement thereafter (12th up-to 60th month).

The most important finding from this analysis is that persistent disease level groups LDA, MDA, and HDA were associated with different 5-year functionality trajectories.

Table 6. Multivariable mixed-effect regression analysis associated groups MDA and HDA with worse 5-year functionality (HAQ) trajectory than LDA group.

HAQ Multivariable Analysis†	**Coefficient‡**	**95% CI**	**p-value**
Group MDA (vs LDA)	+0.273	+0.22 to +0.33	p<0.0001*
Group HDA (vs LDA)	+0.69	+0.63 to +0.75	p<0.0001*
Gender Female (vs Male)	+0.15	+0.1 to +0.19	p=0.0014*
Age (per year)	+0.01	+0.008 to +0.011	p<0.0001*
Disease Duration (per year)	+0.002	+0.001 to +0.005	p=0.165
Time 3-9 Months (vs baseline)	-0.192	-0.16 to -0.225	p<0.0001*
Time 9-15 Months (vs baseline)	-0.234	-0.202 to -0.267	p<0.0001*
Time 15-21 Months (vs baseline)	-0.249	-0.164 to -0.307	p<0.0001*
Time 21-27 Months (vs baseline)	-0.258	-0.217 to -0.282	p<0.0001*
Time 27-33 Months (vs baseline)	-0.31	-0.251 to -0.34	p<0.0001*
Time 33-42 Months (vs baseline)	-0.274	-0.243 to -0.305	p<0.0001*
Time 42-54 Months (vs baseline)	-0.263	-0.232 to -0.294	p<0.0001*
Time 54-60 Months (vs baseline)	-0.318	-0.287 to -0.349	p<0.0001*

Terms: LDA= Persistent Low Disease; MDA= Persistent Medium Disease; HDA= Persistent High Disease; RMSE = Root mean square error; R^2 = R-squared.

† Multivariable (explanatory) analysis efficiency: RMSE=0.37, R^2=0.63

‡ Increase (+) or decrease (-) in HAQ units, associated with category membership or unit increase, in categorical and continuous variable-type, respectively (regression coefficient).

***** Below significance threshold p=0.05.

Multivariate linear mixed effect regression analysis (explanatory association analysis) was also performed within MDA patients (patient subgroups lower-MDA and higher MDA) to analyze variables associated with their longitudinal 5-year course of patient functionality. In multivariable mixed-effect regression analysis (Table 7) the higher-MDA subgroup was associated with worse 5-year functionality trajectory than the lower-MDA subgroup (+0.26 higher HAQ trajectory in higher-MDA, 95% CI +0.17 to +0.36, p<0.0001). Analysis was adjusted for possible confounding effects on gender, age and disease duration. The 5-year functionality was also significantly worse in females than males (+0.12 higher HAQ trajectory in females, 95% CI +0.01 to +0.24, p=0.04) and in patients of older age (+0.009 higher HAQ trajectory per year, 95% CI +0.005 to +0.013, p<0.0001). At 5 years, both subgroups were associated with improved functionality as compared to baseline (HAQ decrease from baseline: -0.38 HAQ, 95% CI -0.463 to -0.297, p<0.0001). Notably, the improvement occurred mainly within the first 12 months of treatment (HAQ decrease from baseline in first 12 months: -0.285, 95% CI -0.373 to -0.196, p<0.0001) showing minor further improvement thereafter (12th up-to 60th month).

Table 7. Multivariable mixed-effect regression associated subgroup higher-MDA with worse 5-year functionality (HAQ) trajectory than lower-MDA subgroup.

HAQ Multivariable Analysis†	**Coefficient‡**	**95% CI**	**p-value**
Group higher-MDA (vs lower-MDA)	+0.26	+0.17 to +0.36	p<0.0001*
Gender Female (vs Male)	+0.12	+0.01 to +0.24	p=0.04*
Age (per year)	+0.009	+0.005 to +0.013	p<0.0001*
Disease Duration (per year)	+0.002	-0.004 to +0.007	p=0.543
Time 3-9 Months (vs baseline)	-0.200	-0.287 to -0.114	p<0.0001*
Time 9-15 Months (vs baseline)	-0.285	-0.373 to -0.196	p<0.0001*
Time 15-21 Months (vs baseline)	-0.291	-0.378 to -0.204	p<0.0001*
Time 21-27 Months (vs baseline)	-0.279	-0.369 to -0.190	p<0.0001*
Time 27-33 Months (vs baseline)	-0.342	-0.444 to -0.240	p<0.0001*
Time 33-42 Months (vs baseline)	-0.308	-0.392 to -0.225	p<0.0001*

Time 42-54 Months (vs baseline)	-0.313	-0.396 to -0.231	p<0.0001*
Time 54-60 Months (vs baseline)	-0.380	-0.463 to -0.297	p<0.0001*

***** Variable is associated with patients' 5-year functionality (HAQ) course (p<0.05 significance threshold).

‡ Regression coefficient that represents increase (+) or decrease (-) in 5-year functionality (HAQ) course associated with the variable (category membership for categorical variable or unit increase for continuous variable).

† Efficiency of multivariable analysis: RMSE (Root mean square error) = 0.366, R^2 (R-squared) = 0.538.

Terms: lower-MDA= Persistent Lower Medium Disease; higher-MDA= Persistent Higher Medium Disease.

The most important finding from this analysis is that MDA group is rather heterogeneous including subgroups lower-MDA and higher-MDA that were associated with different 5-year functionality trajectories.

3.5 Early Predictors for Patient Group Classification

Predictive analysis was performed to assist in patient classification between groups LDA, MDA, and HDA using data from patients' early therapy months (first semester of treatment). Two complementary machine learning models were developed based on binary multivariable logistic regression in order to predict patient classification in the three PDL groups, LDA, MDA, and HDA. The first focused on the classification between LDA and the rest of the groups (MDA and HDA) and the second on the classification between MDA and HDA. Both models were adjusted for possible confounding effects of gender, age, disease duration (binary, true for years<2), previous csDMARDs at baseline (binary, true for count<2) and cumulative bDMARDs (binary, true for count>1) in first therapy semester (treatment months 3-9).

The multivariable logistic regression analysis of the first model (Table 8) classified between LDA and the rest of the groups (MDA and HDA) and yielded that baseline characteristics associated with LDA were male gender (OR 0.38 for female gender, 95% CI 0.17 to 0.85, p=0.02), lower baseline disease activity (OR 0.42 per unit of DAS28, 95% CI 0.27 to 0.65, p=0.001), lower baseline functionality (OR 0.3 per unit of HAQ, 95% CI 0.12 to 0.77, p=0.01), and lower first semester's average disease activity (OR 0.2 per unit of first semester's average DAS28, 95% CI 0.13 to 0.31, p<0.001). The multivariable analysis achieved average predictive efficiency of 90.9% ACC, 67% TPR, 96% TNR, and 92% AUC in its evaluation on 10 independent datasets (test datasets of 10-fold cross-validation process).

Table 8. Multivariable logistic regression analysis to predict classification in LDA compared to the rest of the patient groups (MDA and HDA).

LDA Classification (vs MDA and HDA)	Univariable Analysis			Multivariable Analysis†		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Gender Female (vs Male)	0.27	0.16 to 0.43	p<0.001*	0.38	0.17 to 0.85	p=0.02*
Age (per year)	0.96	0.94 to 0.98	p<0.001	0.99	0.96 to 1.02	p=0.4
Disease Duration <2 Years	2.03	1.17 to 3.53	p=0.029*	1.68	0.7 to 3.99	p=0.24
Previous csDMARDs Baseline Count <2	1.86	1.11 to 3.13	p=0.018*	0.95	0.43 to 2.1	p=0.89
BTs 1st Semester Count >1	0.17	0.05 to 0.56	p=0.003*	1.55	0.38 to 6.39	p=0.54
Anti-TNF Baseline	6.73	1.61 to 28.01	p=0.009*	3.58	0.61 to 21.1	p=0.15
DAS28 Baseline (per unit)	0.24	0.18 to 0.33	p<0.001*	0.42	0.27 to 0.65	p=0.001*
HAQ Baseline (per unit)	0.14	0.08 to 0.26	p<0.001*	0.3	0.12 to 0.77	p=0.01*
DAS28 1st Semester (per unit)	0.16	0.11 to 0.23	p<0.001*	0.2	0.13 to 0.31	p<0.001*
ΔHAQ 1st Semester <0	3.43	1.53 to 7.66	p=0.002*	1.84	0.58 to 5.82	p=0.29

***** Below significance threshold 0.05.

† Multivariable analysis efficiency (CV): ACC=90.9%, TPR=67%, TNR=96%, AUC=92%.

Terms: LDA= Persistent Low Disease; MDA= Persistent Medium Disease; HDA= Persistent High Disease; csDMARDs= Conventional Synthetic Disease-modifying Anti-rheumatic Drugs; BTs= Cumulative bDMARDs; 1st Semester= first therapy semester that begins after the first therapy trimester which has been excluded as treatment adaptation period; HAQ= Health Assessment Questionnaire; ΔHAQ= HAQ Difference of average HAQ score in first semester from baseline HAQ score; CV= 10-Fold Cross-Validation; ACC= Accuracy; TPR= Sensitivity; TNR= Specificity; AUC= Area Under Receiver-Operating-Characteristic Curve.

The multivariable logistic regression analysis of the second model (Table 9) classified between MDA and HDA groups and yielded that baseline characteristics significantly associated with MDA were patients of younger age (OR 1.04 per year, 95% CI 1.01 to 1.07, p=0.003), short disease duration (OR 2.65 for disease duration years<2, 95% CI 1.12 to 6.24, p=0.026), prednisolone initiation at baseline (OR 1.81 for prednisolone initiation, 95% CI 1.05 to 3.13, p=0.033), lower baseline disease activity (OR 0.56 per unit of DAS28, 95% CI 0.41 to 0.76, p<0.001), lower baseline functionality (OR 0.21 per unit of HAQ, 95% CI 0.11 to 0.39, p<0.001) and first semester's disease activity (OR 0.42 per unit of first semester's

average DAS28, 95% CI 0.31 to 0.56, p<0.001), functionality improvement on first semester compared to baseline (OR 2.89 for first semester's average HAQ < baseline HAQ, 95% CI 1.49 to 5.59, p=0.002), and absence of serious adverse events on first semester (OR 0.32 for SAEs count>0, 95% CI 0.1 to 0.99, p=0.047). The multivariable analysis achieved average predictive efficiency of 80.2% ACC, 61% TPR, 90% TNR, and 85% AUC in its evaluation on 10 independent datasets (test datasets of 10-fold cross-validation process).

Table 9. Multivariable logistic regression analysis to predict classification in MDA compared to HDA patient group.

MDA Classification	**Univariable Analysis**			**Multivariable Analysis†**		
(vs HDA)	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Gender Female (vs Male)	0.61	0.34 to 1.1	p=0.1	0.88	0.4 to 1.94	p=0.76
Age (per year)	0.99	0.97 to 1.01	p=0.33	1.04	1.01 to 1.07	p=0.003*
Disease Duration <2 Years	2.11	1.08 to 4.1	p=0.029*	2.65	1.12 to 6.24	p=0.026*
Previous csDMARDs Baseline Count <2	1.89	1.06 to 3.38	p=0.032*	1.43	0.68 to 2.99	p=0.346
BTs 1st Semester Count >1	0.55	0.33 to 0.93	p=0.025*	1.77	0.85 to 3.66	p=0.125
Prednisolone Baseline	2.3	1.5 to 3.55	p<0.001*	1.81	1.05 to 3.13	p=0.033*
DAS28 Baseline (per unit)	0.36	0.28 to 0.46	p<0.001*	0.56	0.41 to 0.76	p<0.001*
HAQ Baseline (per unit)	0.23	0.14 to 0.36	p<0.001*	0.21	0.11 to 0.39	p<0.001*
DAS28 1st Semester (per unit)	0.33	0.25 to 0.42	p<0.001*	0.42	0.31 to 0.56	p<0.001*
ΔHAQ 1st Semester <0	2.24	1.41 to 3.55	p=0.001*	2.89	1.49 to 5.59	p=0.002*
SAEs 1st Semester Count >0	0.47	0.14 to 0.83	p=0.047*	0.32	0.1 to 0.99	p=0.047*

Terms: MDA= Persistent Medium Disease; HDA= Persistent High Disease; csDMARDs= Conventional Synthetic DMARDs; 1st Semester= first therapy semester that begins after the first therapy trimester which has been excluded as treatment adaptation period; BTs= Cumulative Biological DMARDs; HAQ= Health Assessment Questionnaire; ΔHAQ= HAQ Difference of average HAQ score in first semester from baseline HAQ score; SAEs= Serious Adverse Events; CV= 10-Fold Cross-Validation; ACC = Accuracy; TPR = Sensitivity; TNR = Specificity; AUC = Area Under ROC Curve.

† Multivariable analysis efficiency (CV): ACC=80.2%, TPR=90%, TNR=61%, AUC=85%.

***** Below significance threshold 0.05.

3.6 AI Rule Theory for Logical Analysis (Reasoning)

In this section, we describe the results from the development of the Artificial Intelligence (AI) logical rule theory to infer a prognosis for a patient's PDL group (LDA, MDA or HDA). The theory is loaded into a reasoning engine integrated in the CDS-RA system (Figure 3, AI Layer of the CDS-RA system framework) to enable logical analysis (reasoning) and infer meaningful conclusions.

The AI theory rules are presented below. Rules 1-21 (provided below) were described in detail in the methodology section 2.4.5. In this section we also present rules 22-25 that were included in the AI theory based on the results of the predictive analysis which was described in section 3.5. Essentially, rules 22-25 represent the logic of the two machine learning models for patient classification between LDA, MDA, and HDA using patients' early therapy data (first treatment semester). Specifically, the first model focused on patient classification between LDA and the rest of the groups (MDA and HDA) while the second on patient classification between MDA and HDA. Both models were implemented based on binary logistic regression analysis. The models rules represent mathematical formulas in infix notation as abbreviations for the corresponding formulas in prefix notation.

The first machine learning model that classifies between LDA and the rest of groups (MDA and HDA) is expressed by *"rule_22"* and *"rule_23"*. The *"rule_22"* specifies the preconditions that predict LDA as the patient's group (patient is defined by unique identifier Patient_ID) while the *"rule_23"* specifies the preconditions that do <u>not</u> predict the LDA group (note the negation symbol ¬ on the conclusion of *"rule_23"* "$\neg predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA)$") and indicate that the group is either MDA or HDA. In both rules, lines 1-9 derive the patient variables required for classification (as described in section 3.5), lines 10-13 specify the logistic regression analysis that computes a patient score from the variables (line 10-12: the linear regression score computation, line 13: the logistic function transformation of the linear score), and line 14 defines the score threshold for classification.

The second machine learning model that classifies between MDA and HDA is expressed by *"rule_24"* and *"rule_25"*. The *"rule_24"* specifies the preconditions that predict MDA as the patient's group (patient is defined by unique identifier Patient_ID) while the *"rule_25"* specifies the preconditions that predict HDA. In both rules, lines 1-10 derive the patient variables required for classification (as described in section 3.5), lines 11-14 specify the logistic regression analysis that computes a patient score from the variables (line

11-13: the linear regression score computation, line 14: the logistic function transformation of the linear score), and line 15 defines the score threshold for classification.

$\boldsymbol{rule\_1}$: $patient\_has\_4\_TT\_average\_DAS\_in\_range(ID, remission\_or\_low)$
$\Longrightarrow belongs(ID, groupLDA)$

$\boldsymbol{rule\_2}$: $patient\_has\_4\_TT\_average\_DAS\_in\_range(ID, moderate)$
$\Longrightarrow belongs(ID, groupMDA)$

$\boldsymbol{rule\_3}$: $patient\_has\_4\_TT\_average\_DAS\_in\_range\ (ID, high)$
$\Longrightarrow belongs(ID, groupHDA)$

$\boldsymbol{rule\_4}$: $\rightarrow conflicting\_pairs ::$
$belongs(ID, groupHDA), belongs(ID, groupMDA), belongs(ID, groupLDA)$

$\boldsymbol{rule\_5}$: $\rightarrow superior(rule\_3, rule\_1)$

$\boldsymbol{rule\_6}$: $\rightarrow superior(rule\_3, rule\_2)$

$\boldsymbol{rule\_7}$: $\rightarrow superior(rule\_2, rule\_1)$

$\boldsymbol{rule\_8}$: $expert\_clinician\_classification\ (Clinician\_ID, Patient\_ID, Group)$
$\Longrightarrow belongs(Patient\_ID, Group)$

$\boldsymbol{rule\_9}$: $\rightarrow superior(rule\_3, rule\_8)$

$\boldsymbol{rule\_10}$: $\rightarrow superior(rule\_2, rule\_8)$

$\boldsymbol{rule\_11}$: $\rightarrow superior(rule\_1, rule\_8)$

$\boldsymbol{rule\_12}$: $predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA)$
$\Longrightarrow belongs(Patient\_ID, groupLDA)$

$\boldsymbol{rule\_13}$: $\neg\, predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA),$
$predict\_MDA\_vs\_HDA(Patient\_ID, Group)$
$\Longrightarrow belongs(Patient\_ID, Group)$

$\boldsymbol{rule\_14}$: $\rightarrow superior(rule\_1, rule\_12)$

$\boldsymbol{rule\_15}$: $\rightarrow superior(rule\_2, rule\_12)$

$\boldsymbol{rule\_16}$: $\rightarrow superior(rule\_3, rule\_12)$

$\boldsymbol{rule\_17}$: $\rightarrow superior(rule\_1, rule\_13)$

$\boldsymbol{rule\_18}$: $\rightarrow superior(rule\_2, rule\_13)$

$\boldsymbol{rule\_19}$: $\rightarrow superior(rule\_3, rule\_13)$

$\boldsymbol{rule\_20}$: $\rightarrow superior(rule\_8, rule\_12)$

$\boldsymbol{rule\_21}$: $\rightarrow superior(rule\_8, rule\_13)$

$\boldsymbol{rule\_22}$:

1) $female\_gender(Patient\_ID, V1),$
2) $age(Patient\_ID, Age\_Years), V2 = Age\_Years,$

3) $disease\_duration(Patient\_ID, DD\_Years), V3 = (DD\_Years < 2),$

4) $previous\_csDMARDs\_at\_baseline(Patient\_ID, Count1), V4 = (Count1 < 2),$

5) $first\_semester\_bDMARDs(Patient\_ID, Count2), V5 = (Count1 > 1),$

6) $antiTNF\_initiated(Patient\_ID, V6)$, $baseline\_DAS28(Patient\_ID, V7),$

7) $first\_semester\_DAS28\_avg(Patient\_ID, V8),$

8) $baseline\_HAQ(Patient\_ID, HAQ1), V9 = HAQ1,$

9) $first\_semester\_HAQ\_avg(Patient\_ID, HAQ2), V10 = ((HAQ2 - HAQ1) < 0),$

10) $Score\_Linear\_Regression = 9.944 - 0.957 * V1 - 0.011 * V2 + 0.517 * V3$

11) $- 0.056 * V4 + 0.44 * V5 + 1.274 * V6 - 0.862 * V7$

12) $- 1.613 * V8 - 1.193 * V9 + 0.612 * V10$

13) $Score\_Logistic\_Regression = 1/(1 + e\text{^}(-Score\_Linear\_Regression)),$

14) $Score\_Logistic\_Regression \geq 0.5$

15) $\rightarrow predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA)$

***rule*_23**:

1) $female\_gender(Patient\_ID, V1),$

2) $age(Patient\_ID, Age\_Years), V2 = Age\_Years,$

3) $disease\_duration(Patient\_ID, DD\_Years), V3 = (DD\_Years < 2),$

4) $previous\_csDMARDs\_at\_baseline(Patient\_ID, Count1), V4 = (Count1 < 2),$

5) $first\_semester\_bDMARDs(Patient\_ID, Count2), V5 = (Count1 > 1),$

6) $antiTNF\_initiated(Patient\_ID, V6)$, $baseline\_DAS28(Patient\_ID, V7),$

7) $first\_semester\_DAS28\_avg(Patient\_ID, V8),$

8) $baseline\_HAQ(Patient\_ID, HAQ1), V9 = HAQ1,$

9) $first\_semester\_HAQ\_avg(Patient\_ID, HAQ2), V10 = ((HAQ2 - HAQ1) < 0),$

10) $Score\_Linear\_Regression = 9.944 - 0.957 * V1 - 0.011 * V2 + 0.517 * V3$

11) $- 0.056 * V4 + 0.44 * V5 + 1.274 * V6 - 0.862 * V7$

12) $- 1.613 * V8 - 1.193 * V9 + 0.612 * V10$

13) $Score\_Logistic\_Regression = 1/(1 + e\text{^}(-Score\_Linear\_Regression)),$

14) $Score\_Logistic\_Regression < 0.5$

15) $\rightarrow \neg predict\_LDA\_vs\_Rest(Patient\_ID, groupLDA)$

***rule*_24**:

1) $female\_gender(Patient\_ID, V1),$

2) $age(Patient\_ID, Age\_Years), V2 = Age\_Years,$

3) $disease\_duration(Patient\_ID, DD\_Years), V3 = (DD\_Years < 2),$

4) $previous\_csDMARDs\_at\_baseline(Patient\_ID, Count1), V4 = (Count1 < 2),$

5) $first\_semester\_bDMARDs(Patient\_ID, Count2), V5 = (Count1 > 1),$

6) $prednisolone\_initiated(Patient\_ID, V6), baseline\_DAS28(Patient\_ID, V7),$

7) $first\_semester\_DAS28\_avg(Patient\_ID, V8),$

8) $baseline\_HAQ(Patient\_ID, HAQ1), V9 = HAQ1,$

9) $first\_semester\_HAQ\_avg(Patient\_ID, HAQ2), V10 = ((HAQ2 - HAQ1) < 0),$

10) $first\_semester\_SAEs(Patient\_ID, Count3), V11 = (Count3 > 0),$

11) $Score\_Linear\_Regression = 7.336 - 0.123 * V1 + 0.038 * V2 + 0.973 * V3$

12) $+ 0.356 * V4 + 0.57 * V5 + 0.595 * V6 - 0.581 * V7$

13) $- 0.87 * V8 - 1.578 * V9 + 1.06 * V10 - 1.153 * V11$

14) $Score\_Logistic\_Regression = 1/(1 + e\text{^}(-Score\_Linear\_Regression)),$

15) $Score\_Logistic\_Regression \geq 0.5$

16) $\rightarrow predict\_MDA\_vs\_HDA(Patient\_ID, groupMDA)$

***rule*_25**:

1) $female\_gender(Patient\_ID, V1),$

2) $age(Patient\_ID, Age\_Years), V2 = Age\_Years,$

3) $disease\_duration(Patient\_ID, DD\_Years), V3 = (DD\_Years < 2),$

4) $previous\_csDMARDs\_at\_baseline(Patient\_ID, Count1), V4 = (Count1 < 2),$

5) $first\_semester\_bDMARDs(Patient\_ID, Count2), V5 = (Count1 > 1),$

6) $prednisolone\_initiated(Patient\_ID, V6), baseline\_DAS28(Patient\_ID, V7),$

7) $first\_semester\_DAS28\_avg(Patient\_ID, V8),$

8) $baseline\_HAQ(Patient\_ID, HAQ1), V9 = HAQ1,$

9) $first\_semester\_HAQ\_avg(Patient\_ID, HAQ2), V10 = ((HAQ2 - HAQ1) < 0),$

10) $first\_semester\_SAEs(Patient\_ID, Count3), V11 = (Count3 > 0),$

11) $Score\_Linear\_Regression = 7.336 - 0.123 * V1 + 0.038 * V2 + 0.973 * V3$

12) $+ 0.356 * V4 + 0.57 * V5 + 0.595 * V6 - 0.581 * V7$

13) $- 0.87 * V8 - 1.578 * V9 + 1.06 * V10 - 1.153 * V11$

14) $Score\_Logistic\_Regression = 1/(1 + e\text{^}(-Score\_Linear\_Regression)),$

15) $Score\_Logistic\_Regression < 0.5$

16) $\rightarrow predict\_MDA\_vs\_HDA(Patient\_ID, groupHDA)$

The representation of the two machine learning models with logical rules (rules 22-25) was selected to support, (a) interpretability (the models are represented by an interpretable and expressive rule-based logical theory), (b) accessibility (the models are not hard-coded into a machine learning system but rather expressed externally in an easily

accessible AI theory file), (c) configurability (theory rules can be modified easily in case the models are improved with larger patient datasets and the variable weights change), and (d) extendability (additional rules and policies can be added to the theory to express more complex classification models, for example on patient subgroup classification).

3.7 Graphical User Interface of the CDS System

This section presents the graphical user interface (GUI) of the clinical decision-support (CDS) system for rheumatoid arthritis, namely CDS-RA and illustrates the workflow of the system's services for the patient and clinician user modules. The system's services are web-based and mobile-compatible supporting desktop and tablet computers in order to facilitate patient-clinician interaction during therapy.

3.7.1 CDS-RA Graphical User Interface for Patient User Module

The graphical user interface of the patient user module is presented in Figures 10-12. Specifically, Figure 10 presents the login service, Figure 11 illustrates patient questions about overall health state and pain level and Figure 12 depicts two sample questions from (A) the supported heath assessment questionnaire (HAQ) and (B) the quality of life questionnaire EUROQOL.


Rheumatology Clinic PAGNI
Medicine Department University of Crete
Welcome to the application
Rheumatology Questionnaire!
Continue

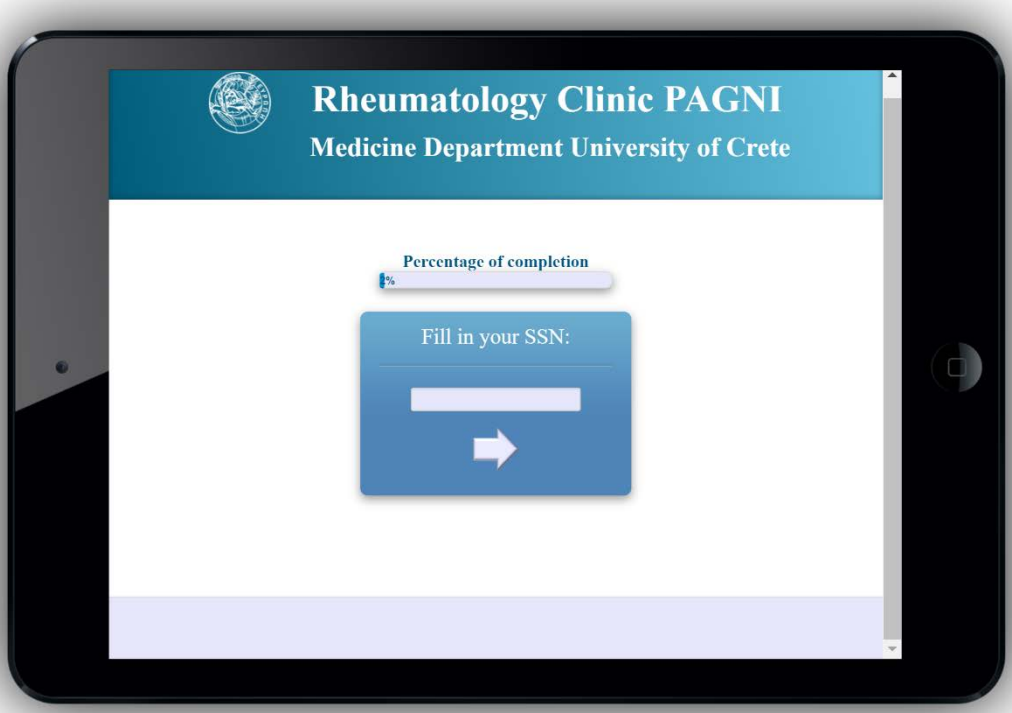
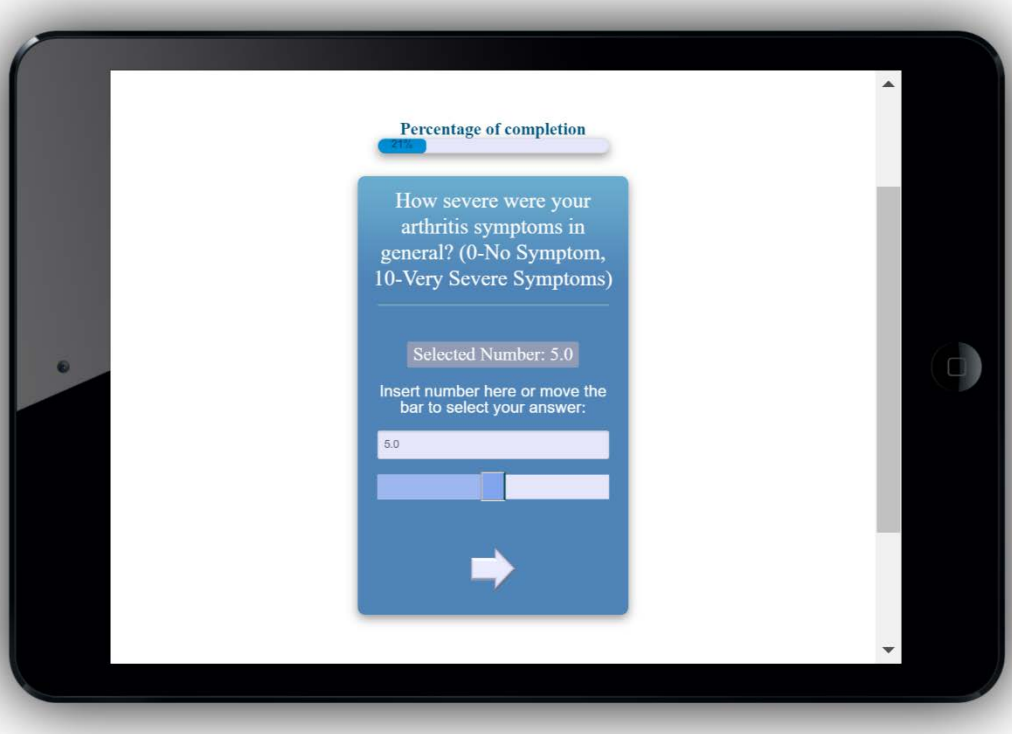


Figure 10. CDS-RA Graphical User Interface (GUI) of Patient Services Login Page.



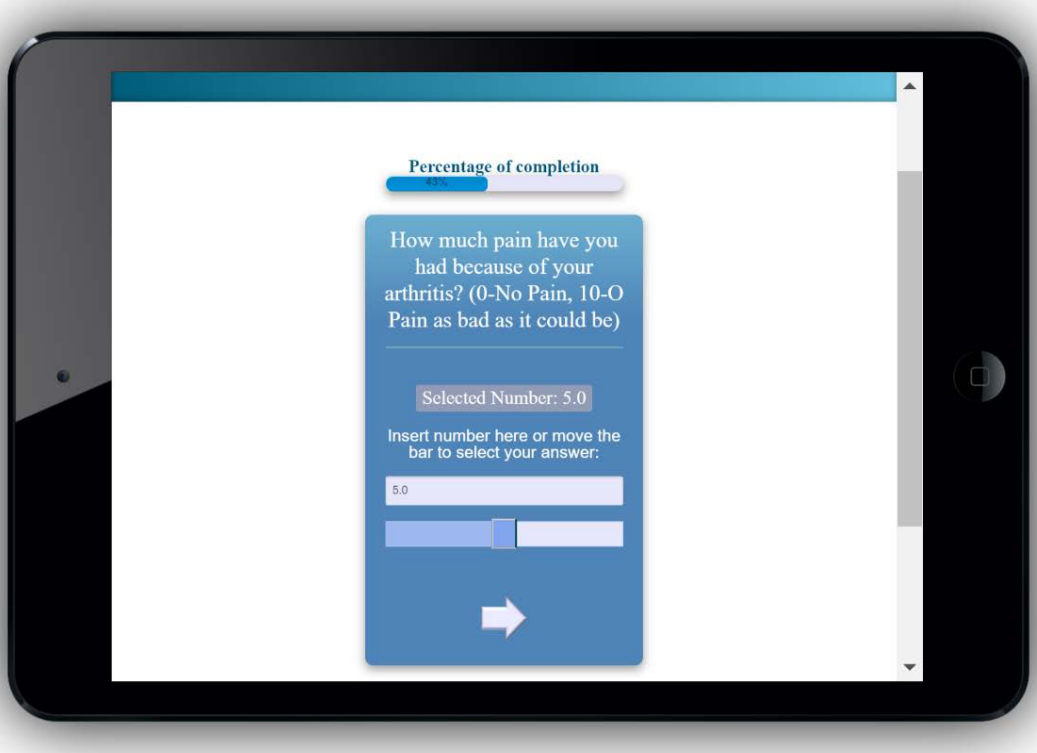
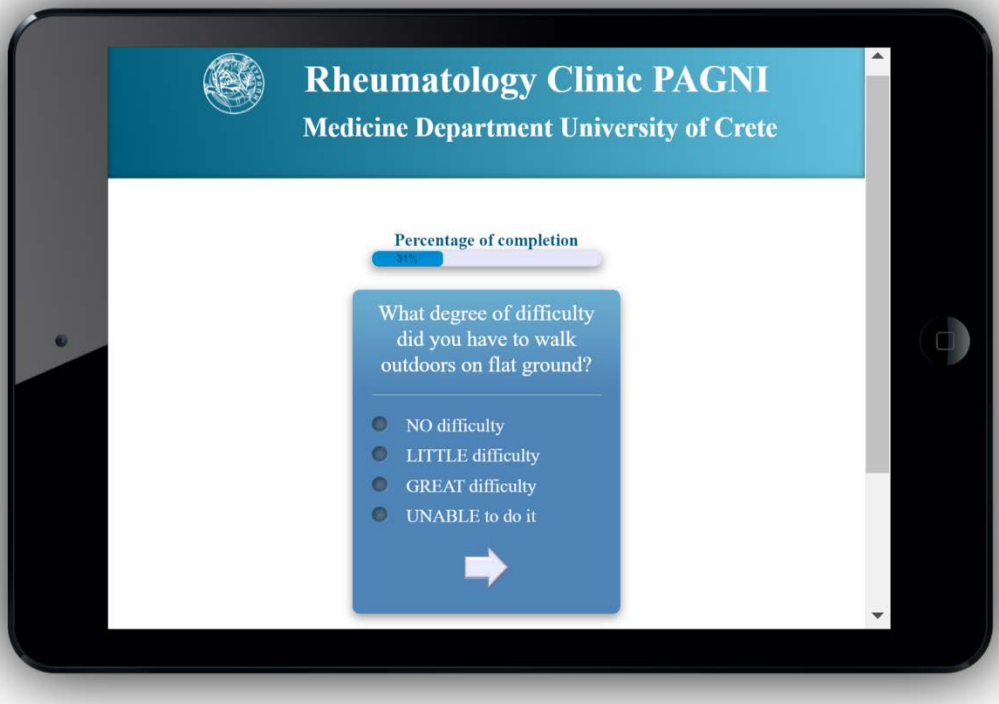


Figure 11. CDS-RA GUI of Patient Questions on Overall Health & Pain Self-Assessment.

A.



B.

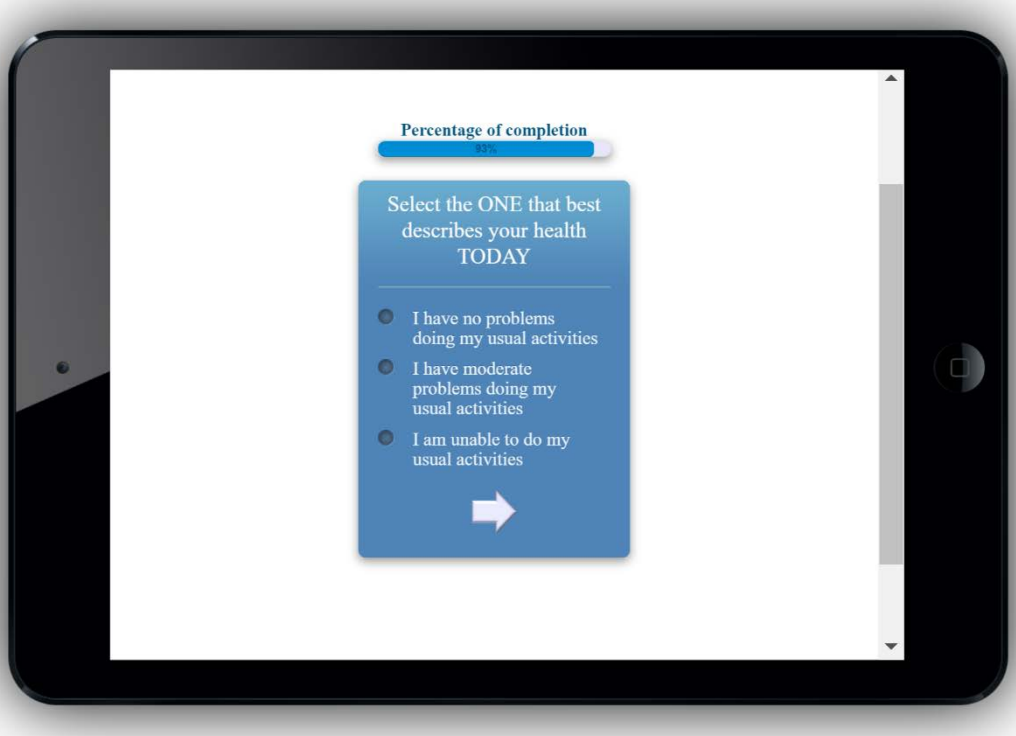


Figure 12. CDS-RA GUI of Questions from the Questionnaires, (**A**) Health Assessment Questionnaire (HAQ) and (**B**) Quality of Life Questionnaire (EUROQOL).

3.7.2 CDS-RA Graphical User Interface for Clinician User Module

The graphical user interface of the clinician user module is presented in Figures 13-21. Specifically, Figure 13 shows (A) the web page to search for a specific patient's data by social security number and (B) the web page presenting 5 services supported on the patient's data.

- Service 1 (Figure 13) leads to the patient's demographic information (Figure 14 A.).
- Service 2 (Figure 13) leads to an illustration of the patient's disease activity (DAS28) trajectory from the clinical assessments (Figure 14 B.).
- Service 3 (Figure 13) is the prognostic service of the patient's persistent disease group (LDA, MDA or HDA). Selection of service 3 (Figure 13) leads to a data form where the clinician is requested to insert additional patient data from the patient's early therapy months (Figure 15 A.). These data are required in one of the three prognostic policies supported by the Artificial Intelligence (AI) engine of the system (section 2.2.4) and

subsequently the prognosis is provided from the AI engine regarding the patient's persistent disease group (Figure 15 B.).

- Service 4 (Figure 13) leads to the presentation of 5-year outcomes associated with persistent disease groups LDA, MDA, and HDA, in terms of functionality (Figure 16 A.) and serious adverse events (Figure 16 B.) trajectories.
- Service 5 supports the ability to select a questionnaire from an interactive list containing all the completed patient questionnaires (Figure 13) that leads to a web page with the following three options depicted in Figure 17, (a) to view patient data summary associated with the completed questionnaire analysis, (b) to add or edit clinical assessment data that are analyzed in the development of the patient data summary, and (c) to add free-text clinical notes associated with the clinical assessment. The selection of the first option to view the patient data summary (Figure 17, Service a.) leads to the summary data table depicted in Figure 18. The selection of the second option to add or edit clinical assessment data (Figure 17, Service b.) leads to the web page in Figure 19 to insert or edit physician's global assessment (VAS Physician), patient erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) and subsequently in Figure 20 to select or edit patient tender and swollen joints. Finally, the selection to add free-text clinical notes (Figure 17, Service c.) leads to the web page in Figure 21 that includes a text-area to edit and save clinical notes.

A.

B.

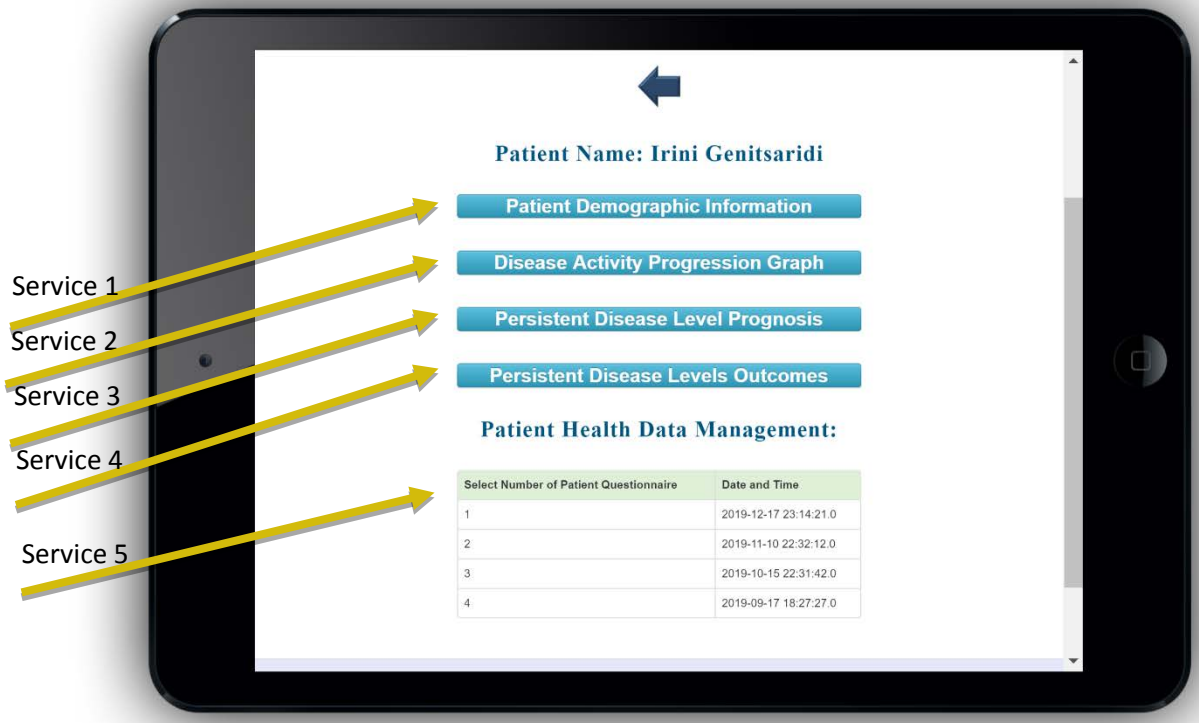
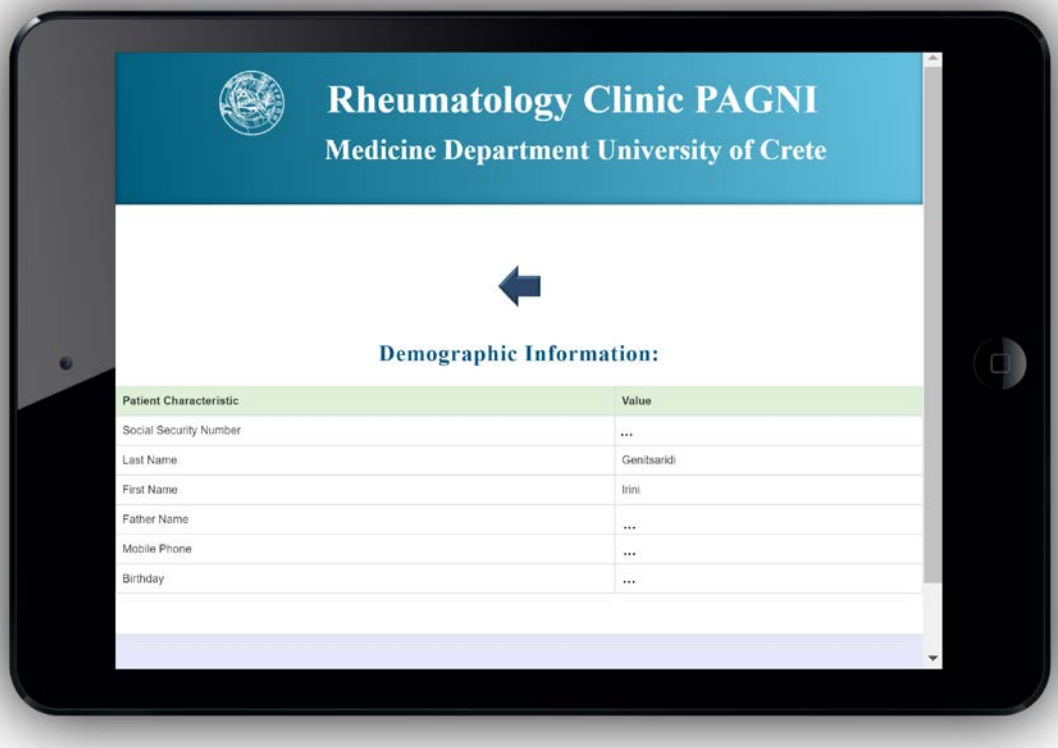


Figure 13. CDS-RA GUI of Clinician Services, (**A**) Specific Patient Data Search by Social Security Number and (**B**) View of Services that are Supported on Patient Data.

A.



B.

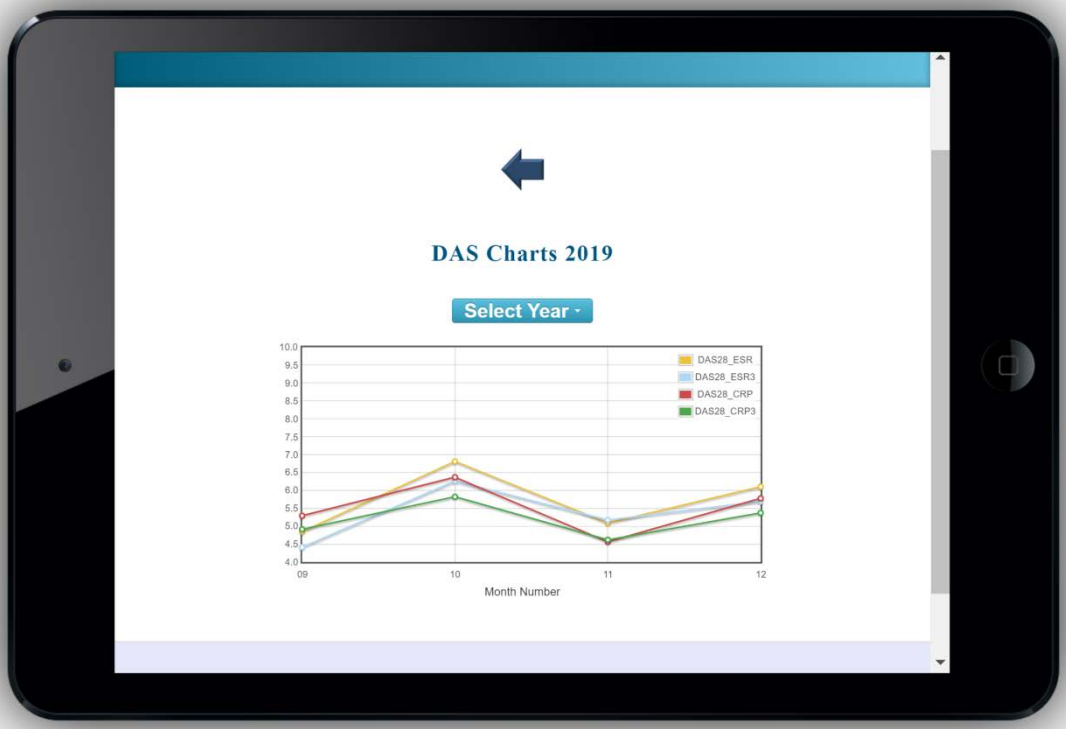
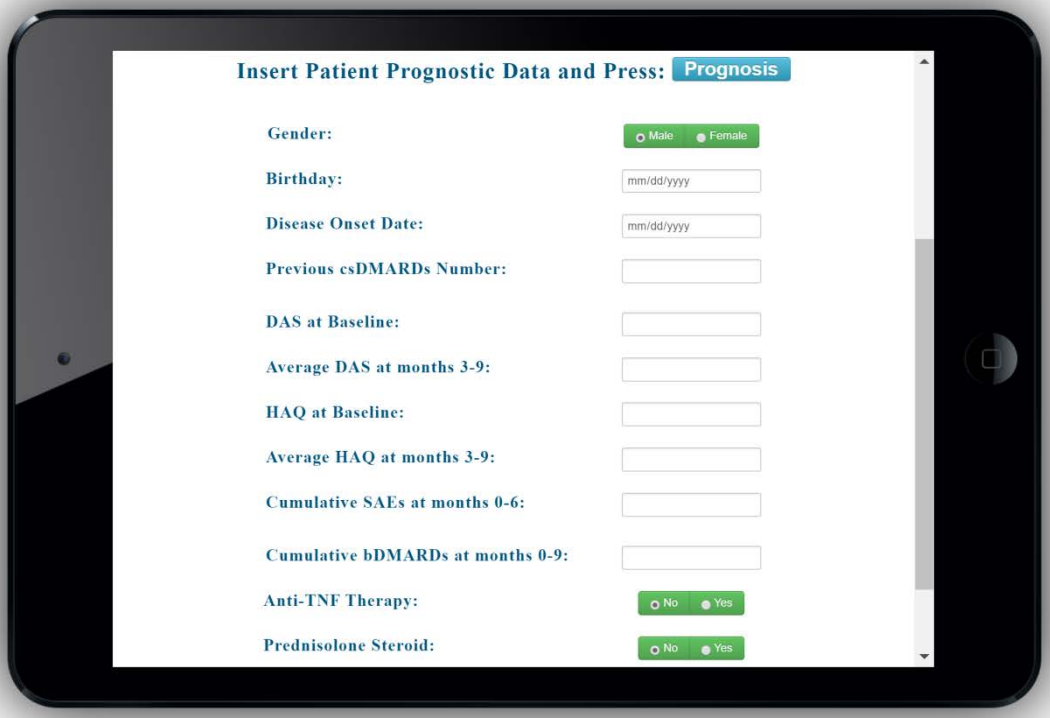


Figure 14. CDS-RA GUI for (**A**) Patient Demographics (**B**) Patient's Disease Activity (DAS28 with ESR and 3-4 variables and DAS28 with CRP and 3-4 variables) Trajectory.

A.



B.

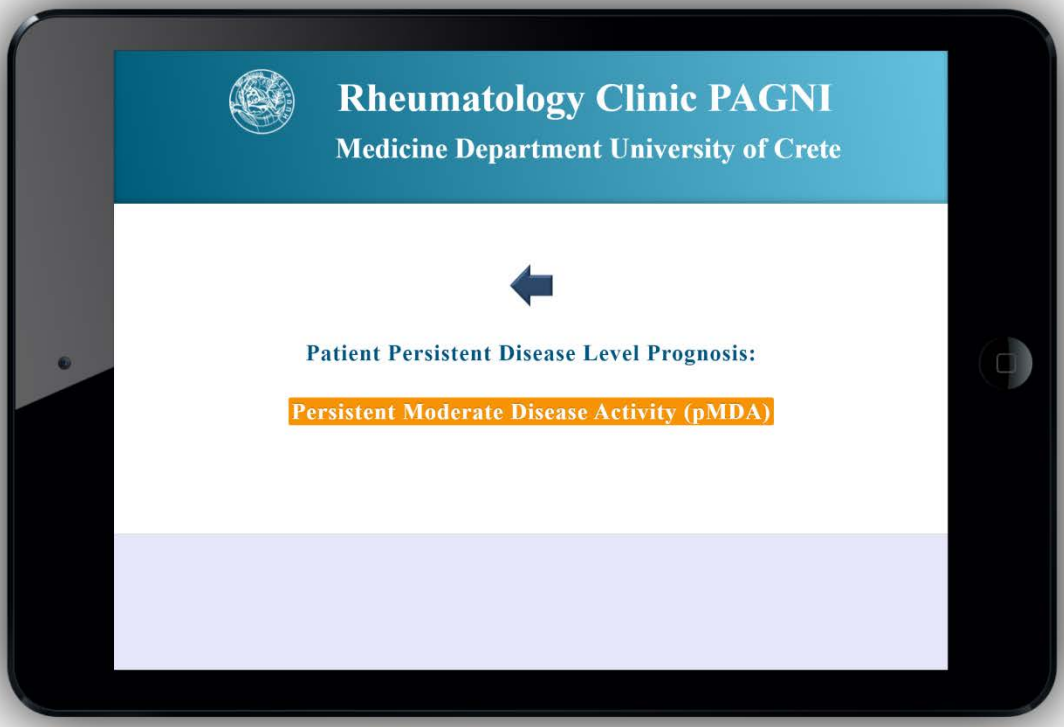
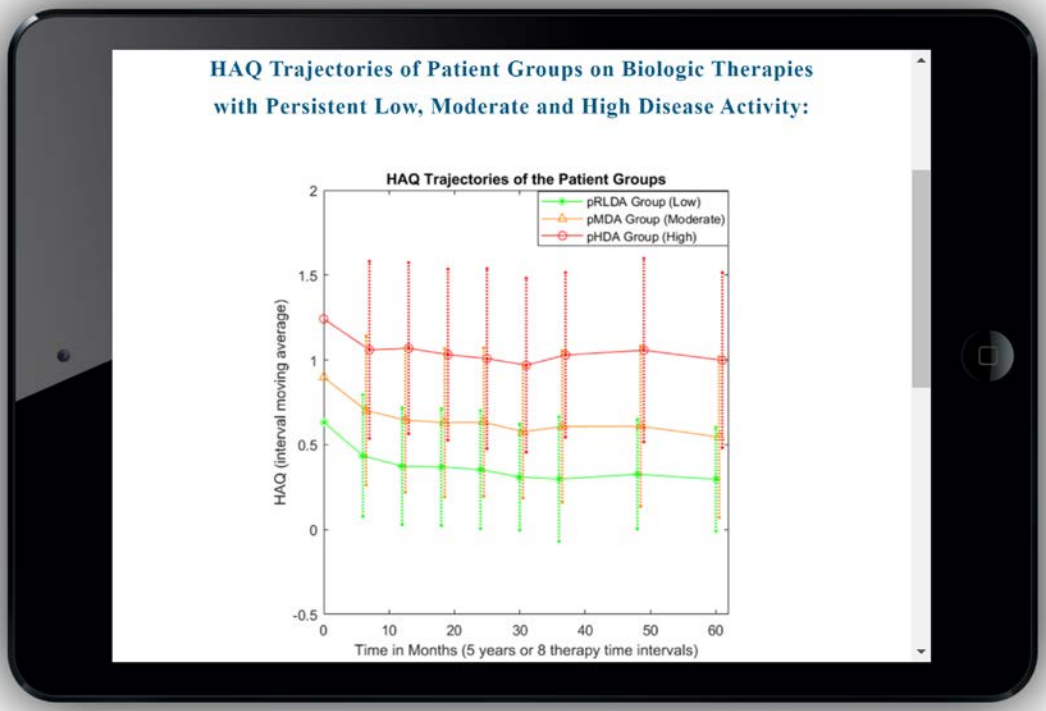


Figure 15. CDS-RA GUI of (**A**) Prognostic AI Service of Patient Persistent Disease Group (LDA, MDA or HDA) Based on Early Therapy Data and (**B**) Prognostic Result View.

A.



B.

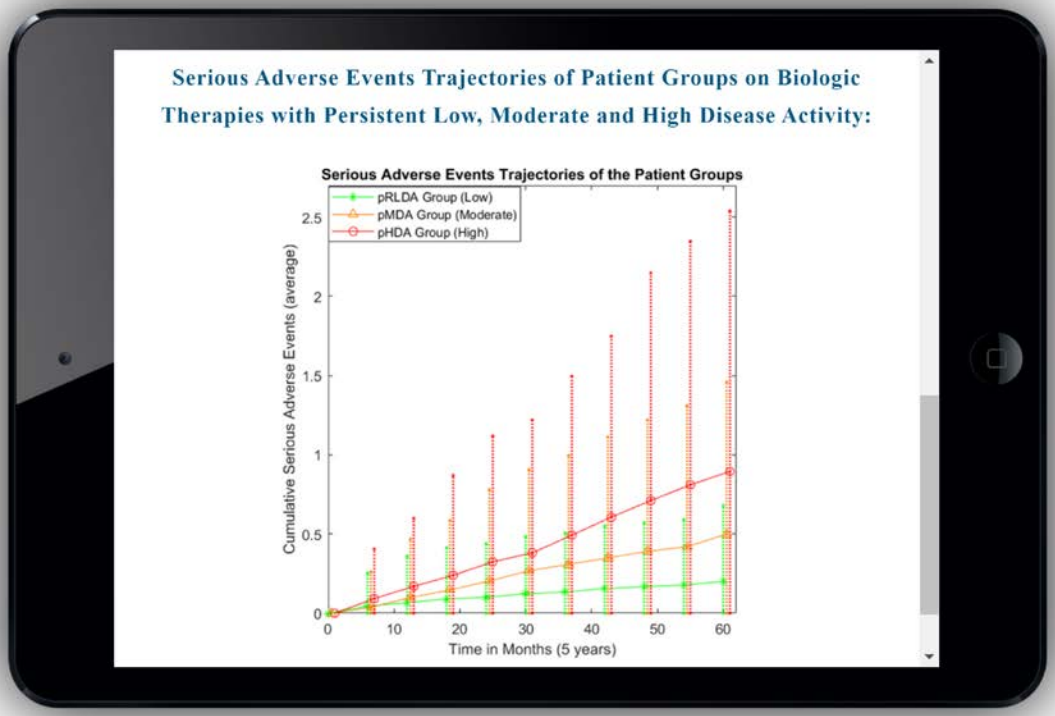


Figure 16. CDS-RA View of 5-Year Trajectories Associated with Groups LDA, MDA, and HDA for (**A**) Functionality (HAQ) and (**B**) Serious Adverse Events.

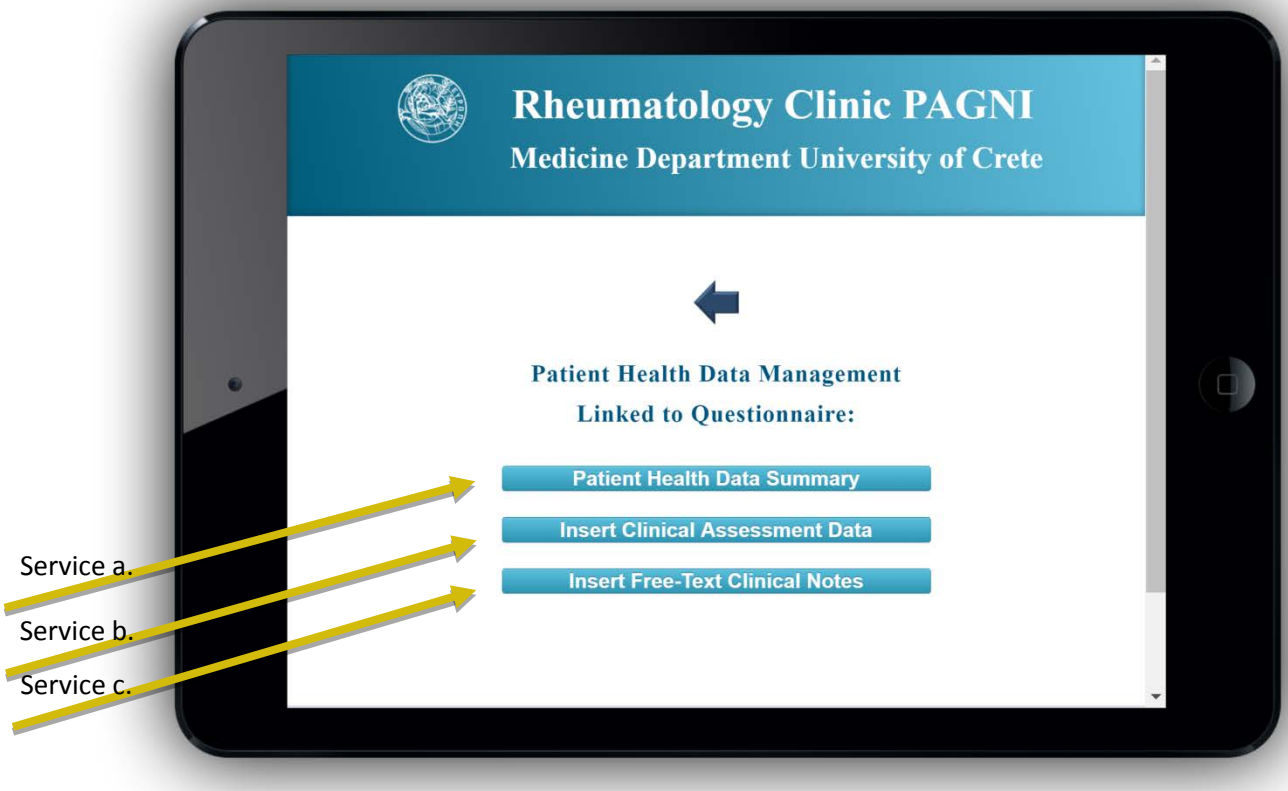


Figure 17. CDS-RA GUI Services to View a Patient Data Summary from a Completed Questionnaire (Service a.) or to Insert Additional Clinical Data (Services b. and c.).

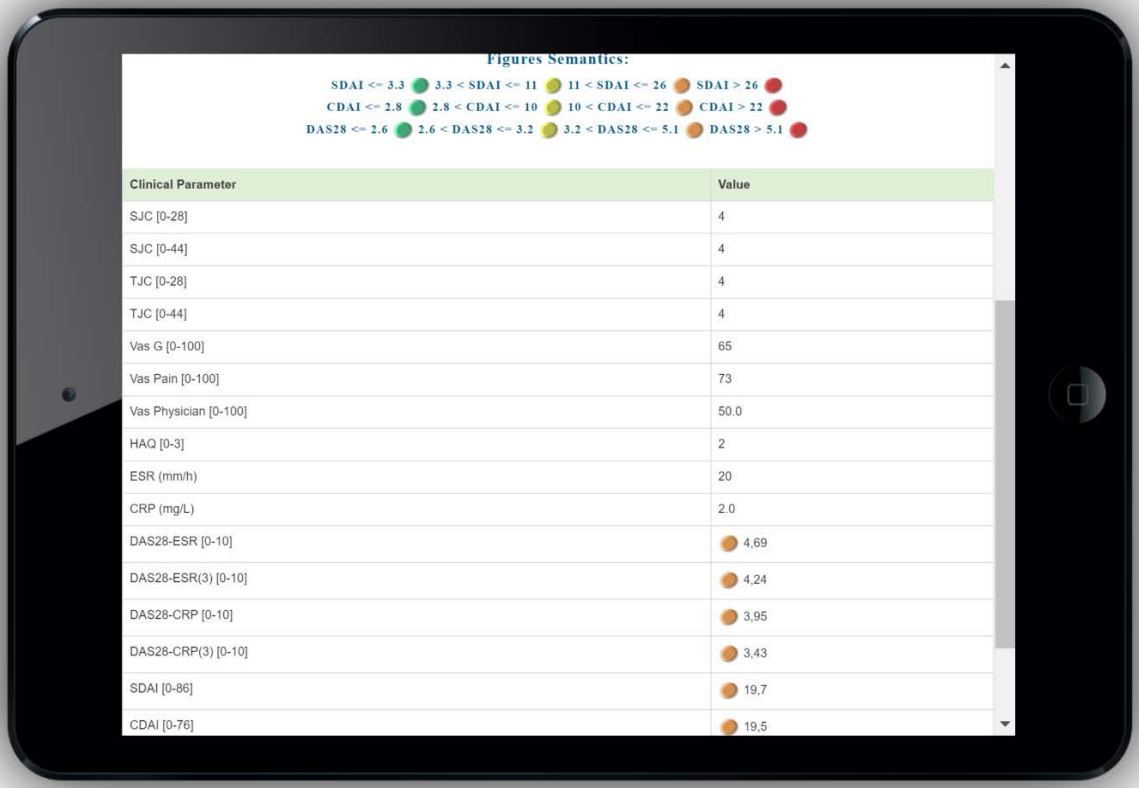


Figure 18. CDS-RA View of a Patient Data Summary Associated with a Completed Questionnaire and Clinical Assessment Data.

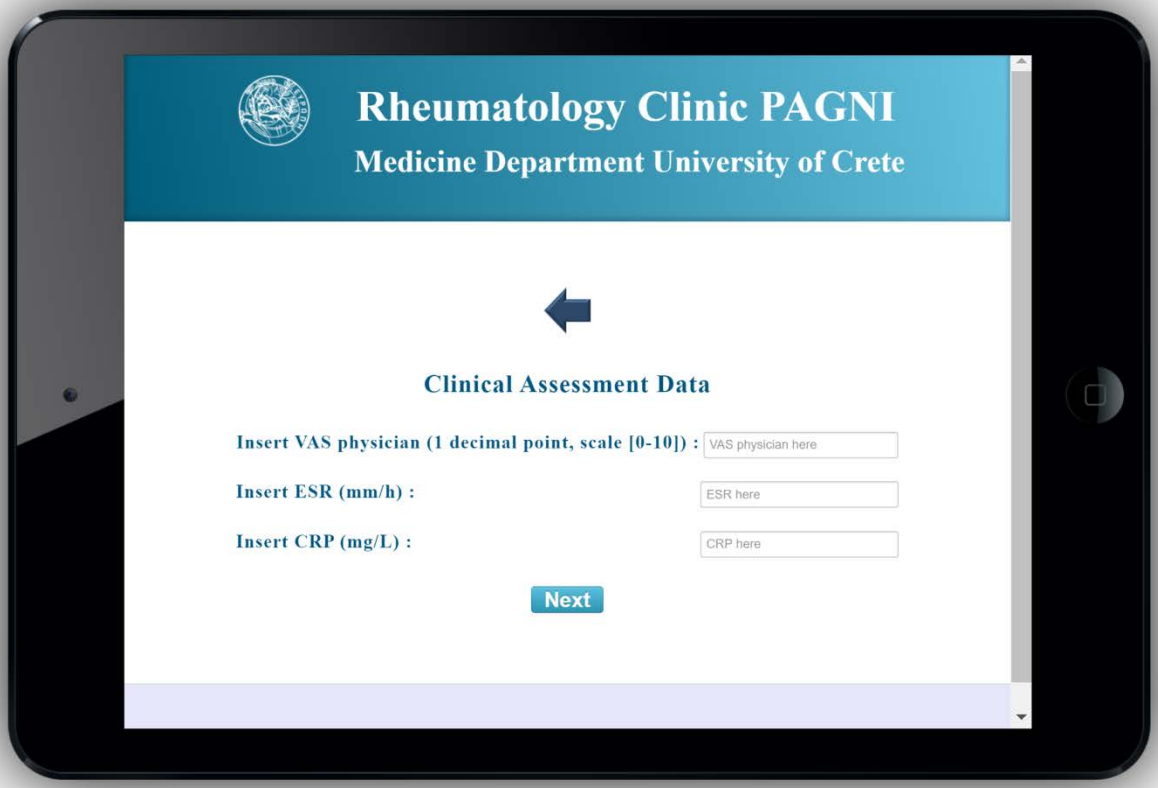


Figure 19. CDS-RA GUI Service to Edit or Add Physician's Global Assessment (VAS Physician), Erythrocyte Sedimentation Rate (ESR) and C-Reactive Protein (CRP).

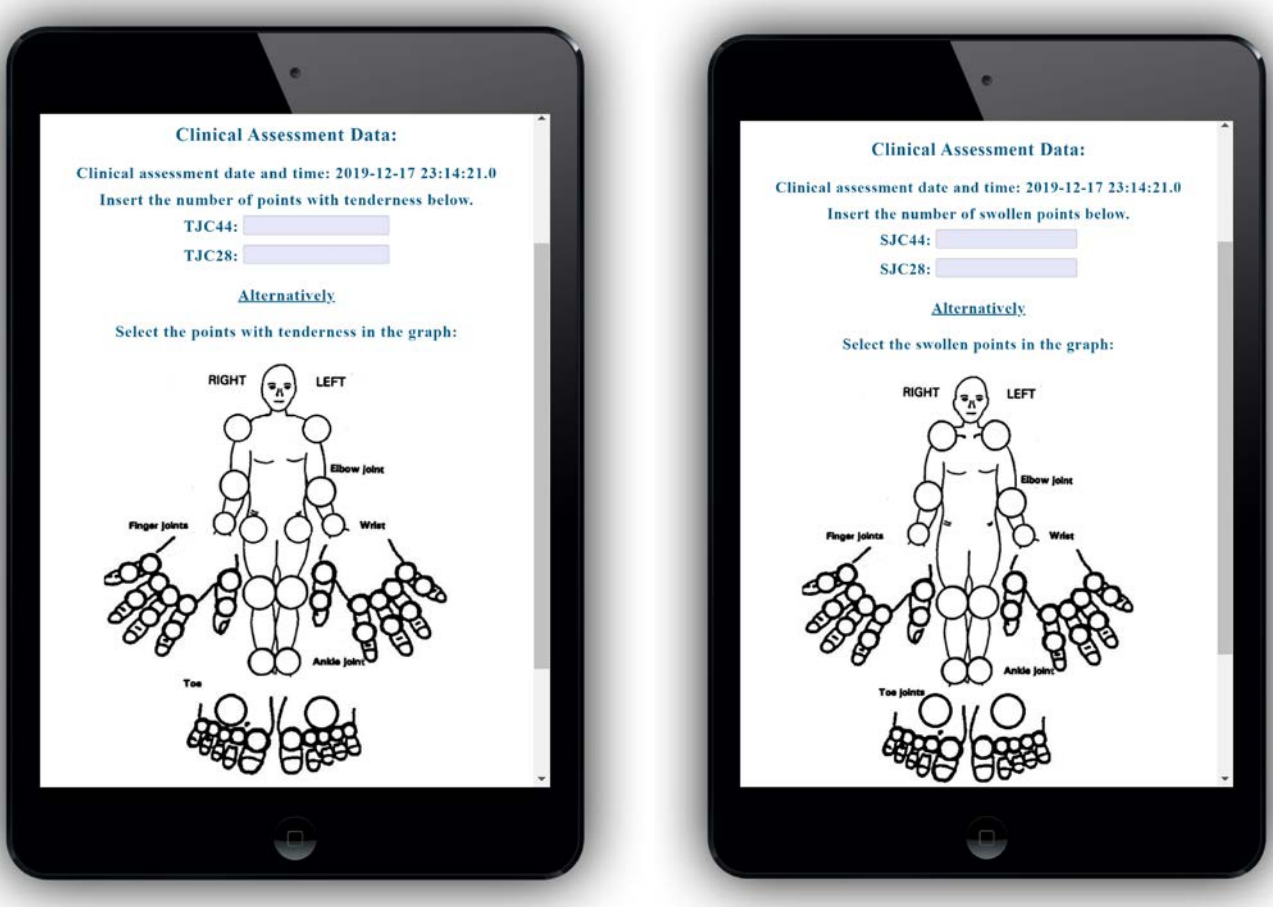
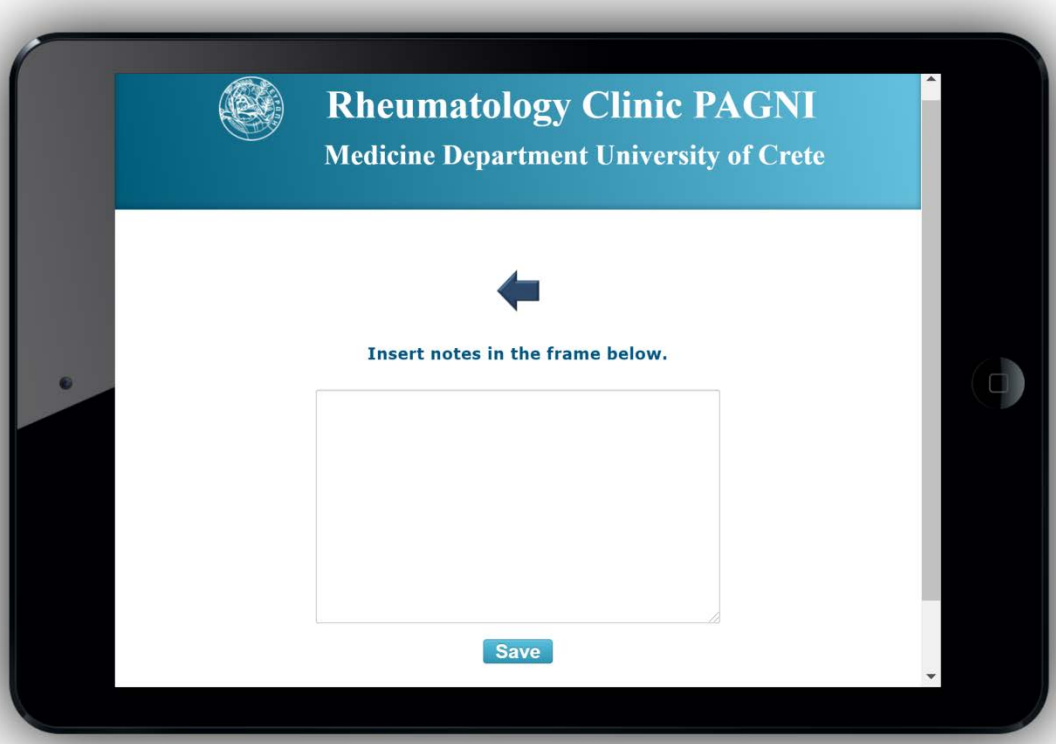


Figure 20. CDS-RA GUI Service to Select Tender and Swollen Patient Joints from an Interactive Image.

A.



B.

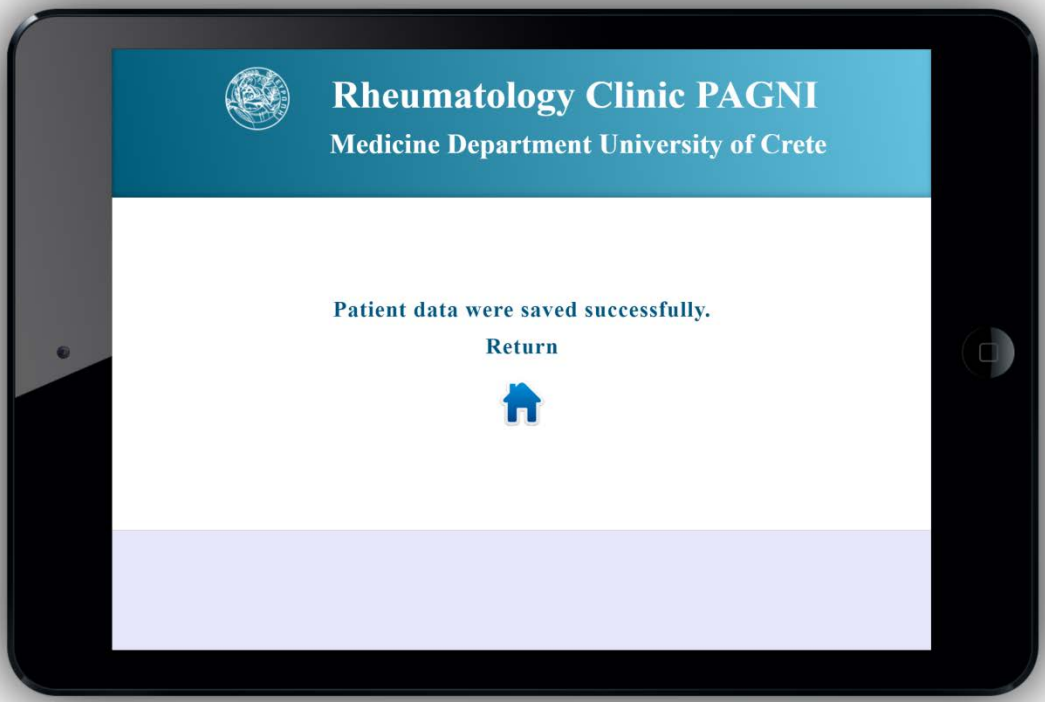


Figure 21. CDS-RA GUI Service to (**A**) Add/Edit and (**B**) Save Clinical Notes.

CHAPTER IV. Discussion & Conclusion

This study focused on the development of a clinical decision support system for rheumatoid arthritis management and long-term prognosis. It was the result of a multidisciplinary effort to support rheumatoid arthritis clinical research domain with artificial intelligence methodologies from the computer science domain. The developed system was based on state-of-the-art AI methods including supervised machine learning, defeasible reasoning and statistical multivariable mixed-effect modeling. In this chapter, we discuss the contribution of this work to RA clinical research as well as the study limitations, possible future work directions, and our final conclusions.

Discussion

The main focus of this study was the application of a variety of artificial intelligence methods in the particular domain of rheumatoid arthritis management and long-term prognosis. Chapter I presented methods for advanced medical data analysis from the following categories, (a) preparatory analyses, (b) association analyses of difference, (c) association analyses of correlation, (d) explanatory association analyses, (e) predictive analyses, and (f) logical analyses. Chapter II described the characteristics of rheumatoid arthritis disease and also related work on the disease management and prognosis with artificial intelligence solutions. In addition, the chapter provided information on literature limitations in the long-term disease prognosis under biologic therapy. Chapter III presented CDS-RA clinical decision support system for rheumatoid arthritis management and long-term prognosis under biologic therapy and analyzed its development methodology and implementation results. In this section, we discuss the contribution and robustness of this work to RA clinical research, the study limitations and the possible future work directions.

1. Study Contribution, Innovation and Robustness

This study presents in detail the CDS-RA system, a seamless AI technological environment focused on clinical decision support for rheumatoid arthritis patients under biologic therapy. All analyses conducted in this study are provided as functional services of CDS-RA system including association, predictive and prognostic analyses. Patients' data required for the CDS-RA system analyses were retrieved from the Greek nationwide multicenter registry

HeRBT (Hellenic Registry of Biologic Therapies) of seven healthcare centers in Greece. Patients were categorized into analysis groups that exhibited persistent low (LDA), medium (MDA) and high (HDA) disease, respectively. Patients in the MDA group were sub-divided into two subgroups for further analysis which exhibited persistent lower medium (lower-MDA) and higher medium (higher-MDA) disease, respectively.

An important finding from the clinical patient distribution into groups was that a substantial proportion exhibited persistent medium disease (MDA) compared to persistent low or high disease (patients: 56% in MDA with 275 vs 17% in LDA with 90 vs 27% in HDA with 142) irrespective of treatment modifications (bDMARD switches and dosage adjustments) in clinical practice. This finding is aligned with other literature study results presenting that approximately 30-50% of patients on biologic drug treatments improve disease activity status with the first treatment, yet they still exhibit moderate disease activity (medium disease level) after the first treatment [85]. It is also aligned with data from other registries showing that only 8.2-21% of RA patients are classified as sustained remission [112-116].

An innovative aspect of the CDS-RA functionality is that it presents information which is scant in literature (Chapter II, section 3) regarding the long-term outcomes of RA patients who exhibit MDA (persistent medium disease) under biologic therapy. MDA is exhibited by a substantial proportion of RA patients under biologics in clinical practice, yet most literature reports focused on RA patients that exhibit sustained remission (under biologics) [112-118]. Thus, the CDS-RA aimed to provide important prognostic information on the long-term outcomes of MDA patients under biologics regarding their functionality (HAQ) course and the occurrence of serious adverse events (SAEs) occurrence during 5 years of therapy. To this end, a number of association analyses on the patients of HeRBT registry, were included in CDS-RA system development process.

A key finding from the association analyses that were performed in the process of the CDS-RA development, was that patient groups exhibiting persistent low (LDA), medium (MDA) and high disease (HDA) under biologic therapy, were associated with different 5-year functionality (HAQ) course. Statistical multivariable mixed-effect modeling was used to model functionality over time. HAQ course was worse in the HDA than the other groups and also it was worse in the MDA than the LDA group (Chapter III, section 3.4). Nevertheless, all groups presented improvement from their baseline functionality status. The results are aligned to those from early-RA cohorts treated on csDMARDs that assessed the cumulative effect of disease activity on RA-related outcomes [95, 96, 119]. Importantly, the present

study focused on patients treated with biologics (bDMARDs) since bDMARDs might exert differential immunomodulatory effects compared to csDMARDs (i.e. TNFi and tocilizumab may inhibit joint destruction effectively even when residual disease activity exists, which is not the case for methotrexate [120-122]).

Another key finding from the association analyses that were performed in the process of the CDS-RA development, was that groups LDA, MDA, and HDA also differed in the occurrence of serious adverse events (occurrence was higher in HDA than in the other groups and also it was higher in the MDA than the LDA group) during 5 years of therapy (Chapter III, section 3.3). The correlation between disease activity level and serious infections has been shown by several cohort studies [123-125]. Nevertheless, only the present study and the CORRONA registry study by Accortt et al. [126] analyzed the "cumulative" disease activity levels, revealing the significant "dose effect" of inflammatory burden on the risk for serious infections. Moreover, data from the Nijmegen early RA inception cohort have shown that time-averaged disease activity burden contributes to the risk of cardiovascular events in RA patients on different background therapies [127, 128]. These findings combined with the finding of higher functional decline in the MDA than in the LDA group, underline the importance of cumulative residual disease activity as an important contributor in RA long-term prognosis.

Another innovative aspect of the present study is that we introduced the metric CTP as representative of the longitudinal course of patient disease activity. CTP represents the cumulative time percentage that DAS28 falls within a specific range during follow-up and irrespective of fluctuations. CTP was also used in the specification of persistent disease level in the present study (DAS28 within a specific range for CTP≥50% of the 5-year follow-up). One of the limitations in the literature is that the majority of short- and even long-term studies evaluate RA-related inflammation cross-sectionally (single time point analysis). However, metrics representative of the longitudinal course of disease activity and its effect over time are considered to provide more valuable information. Other such metrics are the average disease activity from multiple years of treatment (AVG) [96] and the area under the curve of the DAS28 course (AUC) which was associated with both radiographic progression [129] and the risk for cardiovascular diseases (CVD) [127, 128]. The CTP metric compared to the AVG metric is more granular and less prone to outliers in the representation of the patient's disease activity course. In addition, the CTP compared to the AUC metric provides the following advantages, (a) it is more interpretable and flexible to support stricter or looser specifications of persistent disease activity with different CTP thresholds, (b) it can intelligently ignore a

percentage of fluctuations in the patients' disease activity trajectory while restricting a proportion of the treatment course in a specific disease activity level, (c) and most importantly, the AUC metric may not be able to distinguish a persistent moderate disease activity trajectory from one fluctuating equally between low and high disease activity levels since they may exhibit approximately equivalent AUC values.

Sensitivity analyses were also performed in the development of CDS-RA system to evaluate the robustness of the findings regarding the differentiation of groups LDA, MDA, and HDA in their long-term outcomes (data not shown). Specifically, sensitivity analyses focused on a shorter biologic therapy duration including patients with 3-year (instead of 5-year) follow-up and comparing their long-term functionality (HAQ) and serious adverse events (SAEs) at 3 years of therapy between the persistent disease groups LDA, MDA, and HDA. Sensitivity analysis included 766 patients in the persistent disease groups (persistent disease as 3 out of 6 semesters DAS28 activity within groups specified ranges) and the distribution of patient in the groups was approximately the same as in the 5-year analysis. Sensitivity analysis yielded similar results with groups' LDA, MDA, and HDA differing on the 3-year functionality (HAQ). SAEs at 3 years were also different between the HDA and the rest of the groups. Notably, the LDA and MDA did not differ on SAEs at 3 years which indicates that these groups require a larger therapy duration (5 years) to differ on this outcome. Additional analysis of patient inclusion year (3 time intervals from cohort initiation, year<2005, 2005<year<2010, year>2010) did not yield differentiated groups in long-term outcomes and thus it was not included in the analyses.

Another key finding from the association analyses that were performed in the process of CDS-RA development, was that MDA represents a heterogeneous group. Specifically, it was found that patient subgroups in lower medium (lower-MDA) and higher medium (higher-MDA) disease under biologic therapy were associated with different 5-year functionality (HAQ) course. Statistical multivariable mixed-effect modeling was used to model functionality over time. HAQ course was worse in higher-MDA than lower-MDA (Chapter III, section 3.4) and the subgroups also differed in the occurrence of serious adverse events (occurrence was higher in higher-MDA than lower-MDA) during 5 years of therapy (Chapter III, section 3.3). The differentiation between lower-MDA and higher-MDA patient subgroups is an important clinical research finding that provides evidence on the heterogeneity of patients with persistent medium disease (MDA). This finding is in line with other findings from early RA cohorts treated on csDMARDs showing that patients on the lower end of MDA have significantly better outcomes than those on the higher end [96, 130].

The heterogeneity found within MDA patient group can assist T2T strategies to tailor therapeutic strategies for the heterogeneous MDA subgroups in order to improve their adverse long-term prognosis.

Another innovative aspect of the CDS-RA functionality is that it supports a prognostic service able to provide a personalized prognosis for a patient's persistent disease group (LDA, MDA, and HDA) based on various medical evidence sources (Chapter III, sections 2.4.5 and 3.6). This service utilizes a reasoning engine that loads an AI rule-based logical theory to enable logical analysis and infer meaningful conclusions. The AI logical theory is developed in Defeasible Logic and includes three prioritized medical reasoning policies. The first policy is based on patient's long-term follow-up data (if they exist) which may fulfill the classification criteria of groups LDA, MDA, and HDA by their definition, the second policy is based on clinicians' expert opinions (when provided) while the third policy is based on a developed predictive service that utilizes only patient early (first 6-9 months) therapy data. The policies also specify rule priorities to address possible contradictory medical evidence that may arise between them. The AI logical theory is configurable and can be enriched with additional policies from other medical evidence sources.

A key finding from the development of the CDS-RA system predictive service that utilizes only early (first 6-9 therapy months) patient therapy data to classify RA patients under biologic therapy into groups LDA, MDA, and HDA, was the identification of early classification predictors (Chapter III, sections 2.4.4 and 3.5). This service was based on the development of two binary machine learning logistic regression models, the first to predict between LDA and the rest of groups and the second between the MDA and HAD groups. Early predictors associated with patient classification in the LDA compared to the other groups (MDA and HDA) were, male gender, lower baseline disease activity, lower baseline functionality and lower first semester's average disease activity. In addition, early predictors associated with patient classification in MDA compared to HDA were, younger age, short disease duration, prednisolone initiation at baseline, lower baseline disease activity, lower baseline functionality, lower first semester's disease activity, functionality improvement on first semester compared to baseline and absence of serious adverse events on first semester. The predictors effect size (ORs) was also analyzed (Chapter III, section 3.5).

The development of the CDS-RA lead to various important findings that can assist clinical research on RA patients under biologics. CDS-RA functionality is able to support patient classification into a persistent disease level group (LDA, MDA, and HDA) and to present the associated long-term outcomes of this group to the clinician. Other system

functionality enables RA patient data management over time and facilitates the patient-clinician interaction. Overall, the system provides an innovative mobile compatible environment that can provide personalized prognostic information for the long-term outcome of RA patients under biologic therapy.

2. Study Limitations

One of the limitations of this study was the missing data in patient disease activity course. Missing data has been a challenge to manage in the analyses of patient data. Management of missing data has been performed prior to classification of patients into LDA, MDA, and HDA groups. We have performed imputation of missing data to patients missing less or up to 50% of their 5-years disease activity monitoring based on modeling the patients DAS28 course with a minimal mixed-effect regression model that accounted for patients' repeated measurements and individual patient variability. It was decided that 166 patient cases exceeding 50% of missing longitudinal disease activity data would induce an unacceptable amount of uncertainty even if imputed. Thus those patients were excluded from this analysis. Imputation to any missing data in the rest of the cases (537 patients) led to 10% increase of included patients in the groups compared to the non-imputed patient dataset. We acknowledge as a limitation the uncertainty that missing data inevitably introduce in the analyses regardless of the methodology (deletion, imputation or herein hybrid) that is selected to address them. For validation purposes the analyses were repeated in the non-imputed dataset yielding the same results in the differentiation of groups long-term outcomes and longitudinal functionality course.

Another limitation of this study can be considered the merging of patients with sustained remission (DAS28<2.6) and persistent strictly low disease activity ($2.6 \leq \text{DAS28} \leq 3.2$) in the LDA group. The LDA group included 52 patients in persistent remission, 20 in persistent strictly low disease activity, and 18 that fluctuated between remission and strictly low disease activity. This merging into one group was implemented due to the small number of patients achieving low ($\text{DAS28} \leq 3.2$) inflammatory burden for a long duration and also because the present study focused mainly on MDA which is exhibited by a substantial proportion of RA patients treated with biologics in clinical practice. It would be interesting to assess in future studies whether the LDA group is heterogeneous in long term outcomes, as was the case for the MDA group.

3. Study Future Directions

In this section, we discuss the possible future directions of this work. The CDS-RA is a clinical-decision support system that integrates methods and results both from the medical and computer science domain. As these domains evolve, there are numerous future studies that can be implemented to improve the CDS-RA environment.

One possible future direction of this work is to repeat the CDS-RA association and predictive analyses with a larger patient dataset since the HeRBT registry is constantly enriched with new RA cases treated with biologics and followed prospectively in the seven healthcare centers. A larger patient dataset could increase the accuracy of the system's predictive analysis that classifies patients into groups LDA, MDA, and HDA based on their early (first 6-9 months) therapy data. In addition, it could help validate the results of the association analyses with patients having a lower percentage of missing data in their disease activity course. Further analysis could also be performed in a larger patient cohort to compare the long-term prognosis of patients exhibiting persistent strictly low ($2.6 \leq \text{DAS28} \leq 3.2$ for CTP≥50% of the 5-year follow-up) disease and patients exhibiting persistent medium disease ($3.2 < \text{DAS28} \leq 5.1$ for CTP≥50% of the 5-year follow-up) while other analysis could focus on the comparison of patients with remission and those with persistent strictly low disease.

Another future direction of this work is to enrich the system's predictive analysis for early patient classification into groups LDA, MDA, and HDA with additional patient parameters that may increase its accuracy. Clinical, genetic and radiographic patient parameters can be included in the CDS-RA predictive models to evaluate their effect in the models' accuracy. Additional predictive models could also be developed to further classify MDA patients into persistent lower medium (lower-MDA) and higher medium (higher-MDA) disease subgroups. Other predictive models could also focus on the further classification of LDA patients into subgroups exhibiting sustained remission and persistent strictly low disease. Furthermore, predictive models could be developed that focus on a specific follow-up interval from baseline such as the 2nd or 3rd treatment semesters (after the 1st treatment semester since this is implemented by the predictive analysis of the present study and focusing only for patients that do not fulfill the classification criteria of any group). Longitudinal predictive models could also be developed that consider natively the patient's

existing disease activity course. In addition, other machine learning modeling methodologies could also be applied to evaluate their classification accuracy compared the implementation of the present study (logistic regression models).

Another future direction of this work is to enrich the AI rule-based logical theory with additional medical reasoning policies from other medical evidence sources. In addition, prioritization of medical reasoning policies could consider formally with rule specifications, the sources reliability (models accuracy etc.), the policies level of specialization (models including parameter subsets of other models etc.) or any other information from the methods applied that may affect the results certainty.

Additional functionality could also be added in the CDS-RA system to recommend treat-to-target strategies and rank them according to the patient's response. Other functionality could also focus on the analysis of the clinical notes that are recorded by the system in order to automatically associate the free-text reported symptoms and patient information created by the clinician with RA comorbidities using natural language processing (NLP) methods.

CDS-RA future studies should also be directed towards the evaluation of the system in real environments in terms of the system usability and the efficiency of the provided insights in the clinicians everyday practice. A short duration pilot usability study was conducted in the rheumatology clinic of PAGNI hospital using tablet devices with positive reviews from the clinical personnel. This pilot study led to important updates in the graphical user interface to support usability from elderly patients. However, a longer evaluation period with specific measurable user feedback and user behavior metrics tracking would provide valuable information on its efficiency and usability and on areas for future improvements.

Conclusion

This thesis focused on the development of a clinical decision-support system (CDS-RA) to support Rheumatoid Arthritis (RA) management and long-term prognosis under biologic therapy. The CDS-RA utilizes artificial intelligence methods to conduct advanced medical data analyses based on statistical mixed-effect models, machine learning and defeasible reasoning.

A key finding from the analyses is that patients exhibiting persistent low (LDA), medium (MDA) and high (HDA) disease levels differ on their 5-year functionality course (higher persistent disease levels exhibit worse longitudinal HAQ trajectories) and serious adverse events (SAEs) occurrence (higher persistent disease levels exhibit higher SAEs occurrence). Another key finding is that a substantial proportion of RA patients under biologic therapy exhibits MDA and those patients represent a heterogeneous group. Evidence for the MDA group heterogeneity was found in the lower and higher MDA subgroups that differed on long-term outcomes. The analysis of patient 5-year functionality course between the patient (sub)groups was based on statistical multivariable mixed-effect modeling.

An innovative aspect of CDS-RA functionality is the service that enables a personalized prognosis of a patient's long-term outcomes by classifying the patient into a persistent disease level group (LDA, MDA, and HDA) which was associated with specific 5-year functionality course and serious adverse events occurrence. The service utilizes a defeasible reasoning engine to infer meaningful conclusions from a developed AI rule-based logical theory. The theory includes three prioritized medical reasoning policies (a) the first is based on patient's long-term follow-up data (if they exist) which may fulfill the classification criteria of groups LDA, MDA, and HDA by their definition, (b) the second is based on clinicians' expert opinions (when provided), and (c) the third is based on a predictive analysis only from early (first 6-9 months) patient therapy data. Two binary machine learning (logistic

regression) classification models were developed to support the third policy and early predictors of patient group classification were identified.

Overall, the CDS-RA provides a state-of-art AI technological environment with a wide range of functional services that is mobile compatible, supports RA patient data management over time, facilitates patient-clinician interaction and provides personalized prognostic information on the long-term outcome of RA patients under biologic therapy. The innovative functionality integrates seamlessly statistical multivariable mixed-effect modeling, machine learning predictive modeling and defeasible logical reasoning to provide valuable insights during the clinical-decision making process. The CDS-RA is aimed to assist clinicians in the biologic therapy of RA patients in order to improve their outcomes and long-term prognosis.

Appendices

A. Appendix – Bibliography

[1] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*: Malaysia; Pearson Education Limited, 2016.

[2] A. Ramesh, C. Kambhampati, J. R. Monson, and P. Drew, "Artificial intelligence in medicine," *Annals of The Royal College of Surgeons of England,* vol. 86, p. 334, 2004.

[3] N. Menachemi and T. H. Collum, "Benefits and drawbacks of electronic health record systems," *Risk management and healthcare policy*, vol. 4, p. 47, 2011.

[4] I. Genitsaridi, H. Kondylakis, L. Koumakis, K. Marias, and M. Tsiknakis, "Towards intelligent personal health record systems: review, criteria and extensions," *Procedia Computer Science,* vol. 21, pp. 327-334, 2013.

[5] I. Genitsaridi, H. Kondylakis, L. Koumakis, K. Marias, and M. Tsiknakis, "Evaluation of personal health record systems through the lenses of EC research projects," *Computers in biology and medicine,* vol. 59, pp. 175-185, 2015.

[6] M. Ross, W. Wei, and L. Ohno-Machado, ""“Big data” and the electronic health record," *Yearbook of medical informatics,* vol. 23, pp. 97-104, 2014.

[7] F. Liu, C. Weng, and H. Yu, "Natural language processing, electronic health records, and clinical research," in *Clinical Research Informatics*, ed: Springer, 2012, pp. 293-310.

[8] M. A. Musen, B. Middleton, and R. A. Greenes, "Clinical decision-support systems," in *Biomedical informatics*, ed: Springer, 2014, pp. 643-674.

[9] S. Piri, D. Delen, T. Liu, and H. M. Zolbanin, "A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble," *Decision Support Systems,* vol. 101, pp. 12-27, 2017.

[10] R. C. Kessler, "The potential of predictive analytics to provide clinical decision support in depression treatment planning," *Current opinion in psychiatry*, vol. 31, pp. 32-39, 2018.

[11] J.-H. Lin and P. J. Haug, "Data preparation framework for preprocessing clinical data in data mining," in *AMIA Annual Symposium Proceedings*, 2006, p. 489.

[12] R. Mlynarski, G. Ilczuk, A. Wakulicz-Deja, and W. Kargul, "A new method of data preparation for cardiological decision support," in *2006 Computers in Cardiology*, 2006, pp. 273-276.

[13] E. Teichmann, E. Demir, and T. Chaussalet, "Data preparation for clinical data mining to identify patients at risk of readmission," in *2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2010, pp. 184-189.

[14] J. Van den Broeck, S. A. Cunningham, R. Eeckels, and K. Herbst, "Data cleaning: detecting, diagnosing, and editing data abnormalities," *PLoS medicine*, vol. 2, p. e267, 2005.
[15] T. Miyagawa, N. Nishida, J. Ohashi, R. Kimura, A. Fujimoto, M. Kawashima, *et al.*, "Appropriate data cleaning methods for genome-wide association study," *Journal of human genetics*, vol. 53, p. 886, 2008.
[16] R. V. Pandey, S. Pabinger, A. Kriegner, and A. Weinhäusel, "ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research," *BMC bioinformatics*, vol. 17, p. 56, 2016.
[17] H. H. Mohamed, T. L. Kheng, C. Collin, and O. S. Lee, "E-CLEAN: A data cleaning framework for patient data," in *2011 First International Conference on Informatics and Computational Intelligence*, 2011, pp. 63-68.
[18] J. A. Sterne, I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, *et al.*, "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls," *Bmj*, vol. 338, p. b2393, 2009.
[19] D. J. Stekhoven and P. Bühlmann, "MissForest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, pp. 112-118, 2011.
[20] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, "Feature extraction and recognition of ictal EEG using EMD and SVM," *Computers in biology and medicine*, vol. 43, pp. 807-816, 2013.
[21] M. Crescenzi and A. Giuliani, "The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data," *Febs Letters*, vol. 507, pp. 114-118, 2001.
[22] D. Walsh and L. Rybicki, "Symptom clustering in advanced cancer," *Supportive care in cancer*, vol. 14, pp. 831-836, 2006.
[23] T. T. Wu, Y. F. Chen, T. Hastie, E. Sobel, and K. Lange, "Genome-wide association analysis by lasso penalized logistic regression," *Bioinformatics*, vol. 25, pp. 714-721, 2009.
[24] K. Juhlin, X. Ye, K. Star, and G. N. Norén, "Outlier removal to uncover patterns in adverse drug reaction surveillance–a simple unmasking strategy," *Pharmacoepidemiology and drug safety*, vol. 22, pp. 1119-1129, 2013.
[25] X.-F. Feng, L.-C. Yang, L.-Z. Tan, and Y.-G. Li, "Risk factor analysis of device-related infections: value of re-sampling method on the real-world imbalanced dataset," *BMC medical informatics and decision making*, vol. 19, p. 185, 2019.
[26] P. H. Rezvan, K. J. Lee, and J. A. Simpson, "The rise of multiple imputation: a review of the reporting and implementation of the method in medical research," *BMC medical research methodology*, vol. 15, p. 30, 2015.
[27] A. K. Waljee, A. Mukherjee, A. G. Singal, Y. Zhang, J. Warren, U. Balis, *et al.*, "Comparison of imputation methods for missing laboratory data in medicine," *BMJ open*, vol. 3, p. e002847, 2013.
[28] J. Cheng, L. J. Edwards, M. M. Maldonado-Molina, K. A. Komro, and K. E. Muller, "Real longitudinal data analysis for real people: building a good enough mixed model," *Statistics in medicine*, vol. 29, pp. 504-520, 2010.
[29] A. Cnaan, N. M. Laird, and P. Slasor, "Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data," *Statistics in medicine*, vol. 16, pp. 2349-2380, 1997.
[30] D. Zhang, L. Zou, X. Zhou, and F. He, "Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer," *IEEE Access*, vol. 6, pp. 28936-28944, 2018.

[31] L. M. De Wit, A. Van Straten, M. Van Herten, B. W. Penninx, and P. Cuijpers, "Depression and body mass index, a u-shaped association," *BMC public health*, vol. 9, p. 14, 2009.
[32] S. Kara and F. Dirgenali, "A system to diagnose atherosclerosis via wavelet transforms, principal component analysis and artificial neural networks," *Expert Systems with Applications*, vol. 32, pp. 632-640, 2007.
[33] A. Wang and E. A. Gehan, "Gene selection for microarray data analysis using principal component analysis," *Statistics in medicine*, vol. 24, pp. 2069-2087, 2005.
[34] P. D'haeseleer, "How does gene expression clustering work?," *Nature biotechnology*, vol. 23, p. 1499, 2005.
[35] C. Samieri, M.-A. Jutand, C. Féart, L. Capuron, L. Letenneur, and P. Barberger-Gateau, "Dietary patterns derived by hybrid clustering method in older people: association with cognition, mood, and self-rated health," *Journal of the American Dietetic Association*, vol. 108, pp. 1461-1471, 2008.
[36] P. Corraini, M. Olsen, L. Pedersen, O. M. Dekkers, and J. P. Vandenbroucke, "Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators," *Clinical epidemiology*, vol. 9, p. 331, 2017.
[37] C.-T. Wu, C. O. Borel, M.-S. Lee, J.-C. Yu, H.-S. Liou, H.-D. Yi, *et al.*, "The interaction effect of perioperative cotreatment with dextromethorphan and intravenous lidocaine on pain relief and recovery of bowel function after laparoscopic cholecystectomy," *Anesthesia & Analgesia*, vol. 100, pp. 448-453, 2005.
[38] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200-1205.
[39] T. Park and G. Casella, "The bayesian lasso," *Journal of the American Statistical Association*, vol. 103, pp. 681-686, 2008.
[40] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Statistics in medicine*, vol. 16, pp. 385-395, 1997.
[41] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, pp. 16-28, 2014.
[42] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate research*, vol. 30, pp. 79-82, 2005.
[43] W. Li and D. R. Nyholt, "Marker selection by Akaike information criterion and Bayesian information criterion," *Genetic Epidemiology*, vol. 21, pp. S272-S277, 2001.
[44] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, pp. 461-464, 1978.
[45] T. M. Cover and J. A. Thomas, *Elements of information theory*: John Wiley & Sons, 2012.
[46] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 1226-1238, 2005.
[47] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, ed: Springer, 2009, pp. 1-4.
[48] L. D. Schroeder, D. L. Sjoquist, and P. E. Stephan, *Understanding regression analysis: An introductory guide* vol. 57: Sage Publications, 2016.
[49] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *bioinformatics*, vol. 23, pp. 2507-2517, 2007.

[50] W. Li, W. Mo, X. Zhang, J. J. Squiers, Y. Lu, E. W. Sellke, *et al.*, "Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging," *Journal of biomedical optics,* vol. 20, p. 121305, 2015.
[51] A. Lourenço, H. Silva, and C. Carreiras, "Outlier detection in non-intrusive ECG biometric system," in *International Conference Image Analysis and Recognition*, 2013, pp. 43-52.
[52] M. Kuklisova-Murgasova, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, and J. A. Schnabel, "Reconstruction of fetal brain MRI with intensity matching and complete outlier removal," *Medical image analysis,* vol. 16, pp. 1550-1564, 2012.
[53] W. Han, Z. Huang, S. Li, and Y. Jia, "Distribution-sensitive unbalanced data oversampling method for medical diagnosis," *Journal of medical Systems,* vol. 43, p. 39, 2019.
[54] J. Li, S. Fong, Y. Sung, K. Cho, R. Wong, and K. K. Wong, "Adaptive swarm cluster-based dynamic multi-objective synthetic minority oversampling technique algorithm for tackling binary imbalanced datasets in biomedical data classification," *BioData mining,* vol. 9, p. 37, 2016.
[55] J. Zhang, L. Chen, and F. Abid, "Prediction of Breast Cancer from Imbalance Respect Using Cluster-Based Undersampling Method," *Journal of Healthcare Engineering,* vol. 2019, 2019.
[56] A. Shintani, "Primer of statistics in dental research: Part II," *Journal of prosthodontic research,* vol. 58, pp. 85-91, 2014.
[57] D. W. Byrne and Byrne, *Publishing your medical research paper: what they don't teach in medical school*: Williams & Wilkins London, UK, 1998.
[58] A. Petrie and C. Sabin, *Medical statistics at a glance*: John Wiley & Sons, 2019.
[59] E. Alpaydin, *Introduction to machine learning*: MIT press, 2020.
[60] T. Hastie and R. Tibshirani, "Generalized additive models for medical research," *Statistical methods in medical research,* vol. 4, pp. 187-196, 1995.
[61] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.
[62] H. Pang, A. Lin, M. Holford, B. E. Enerson, B. Lu, M. P. Lawton, *et al.*, "Pathway analysis using random forests classification and regression," *Bioinformatics,* vol. 22, pp. 2028-2036, 2006.
[63] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
[64] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging,* vol. 3, p. 034501, 2016.
[65] E. Alickovic and A. Subasi, "Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier," *Journal of medical systems,* vol. 40, p. 108, 2016.
[66] M. Frize, C. M. Ennett, M. Stevenson, and H. C. Trigg, "Clinical decision support systems for intensive care units: using artificial neural networks," *Medical engineering & physics,* vol. 23, pp. 217-225, 2001.
[67] P. J. Lisboa and A. F. Taktak, "The use of artificial neural networks in decision support in cancer: a systematic review," *Neural networks,* vol. 19, pp. 408-415, 2006.
[68] J. J. Forsström and K. J. Dalton, "Artificial neural networks for decision support in clinical medicine," *Annals of medicine,* vol. 27, pp. 509-517, 1995.
[69] J.-H. Eom, S.-C. Kim, and B.-T. Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction," *Expert Systems with Applications,* vol. 34, pp. 2465-2479, 2008.

[70] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, pp. 111-133, 1974.
[71] J.-H. Kim, "Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap," *Computational statistics & data analysis*, vol. 53, pp. 3735-3745, 2009.
[72] J. Reichertz, "Induction, deduction," *The Sage Handbook of Qualitative Data Analysis, Sage, London*, p. 123, 2013.
[73] R. Reiter, "Nonmonotonic reasoning," in *Exploring artificial intelligence*, ed: Elsevier, 1988, pp. 439-481.
[74] N. Lachiche, "Abduction and induction from a non-monotonic reasoning perspective," in *Abduction and Induction*, ed: Springer, 2000, pp. 107-116.
[75] D. Nute, "Defeasible reasoning: a philosophical analysis in prolog," in *Aspects of artificial intelligence*, ed: Springer, 1988, pp. 251-288.
[76] D. Nute, "Defeasible logic," in *International Conference on Applications of Prolog*, 2001, pp. 151-169.
[77] M. J. Maher, A. Rock, G. Antoniou, D. Billington, and T. Miller, "Efficient defeasible reasoning systems," *International Journal on Artificial Intelligence Tools*, vol. 10, pp. 483-501, 2001.
[78] D. Aletaha, J. Funovits, and J. S. Smolen, "The importance of reporting disease activity states in rheumatoid arthritis clinical trials," *Arthritis Rheum*, vol. 58, pp. 2622-31, Sep 2008.
[79] J. S. Smolen, F. C. Breedveld, G. R. Burmester, V. Bykerk, M. Dougados, P. Emery, *et al.*, "Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force," *Ann Rheum Dis*, vol. 75, pp. 3-15, Jan 2016.
[80] M. Prevoo, M. A. Van'T Hof, H. Kuper, M. Van Leeuwen, L. Van De Putte, and P. Van Riel, "Modified disease activity scores that include twenty-eight- joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis," *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, vol. 38, pp. 44-48, 1995.
[81] J. S. Smolen, R. Landewe, J. Bijlsma, G. Burmester, K. Chatzidionysiou, M. Dougados, *et al.*, "EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update," *Ann Rheum Dis*, vol. 76, pp. 960-977, Mar 06 2017.
[82] D. Aletaha and J. Smolen, "The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care," *Best practice & research. Clinical rheumatology*, vol. 21, pp. 663-738, 2007.
[83] M. Hetland, I. Christensen, U. Tarp, L. Dreyer, A. Hansen, I. Hansen, *et al.*, "Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: results from eight years of surveillance of clinical practice in the nationwide Danish DANBIO registry," *Arthritis Rheum*, vol. 62, pp. 22-54, 2010.
[84] J. D. Greenberg, G. Reed, D. Decktor, L. Harrold, D. Furst, A. Gibofsky, *et al.*, "A comparative effectiveness study of adalimumab, etanercept and infliximab in biologically naive and switched rheumatoid arthritis patients: results from the US CORRONA registry," *Ann Rheum Dis*, vol. 71, pp. 1134-42, Jul 2012.
[85] I. Flouri, T. E. Markatseli, P. V. Voulgari, K. A. Boki, I. Papadopoulos, L. Settas, *et al.*, "Comparative effectiveness and survival of infliximab, adalimumab, and etanercept for rheumatoid arthritis patients in the Hellenic Registry of Biologics: Low

rates of remission and 5-year drug survival," *Semin Arthritis Rheum,* vol. 43, pp. 447-57, Feb 2014.

[86] C. Gabay, M. Riek, A. Scherer, and A. Finckh, "Effectiveness of biologic DMARDs in monotherapy versus in combination with synthetic DMARDs in rheumatoid arthritis: data from the Swiss Clinical Quality Management Registry," *Rheumatology (Oxford),* vol. 54, pp. 1664-72, Sep 2015.

[87] H. Forsblad-d'Elia, K. Bengtsson, L. E. Kristensen, and L. T. Jacobsson, "Drug adherence, response and predictors thereof for tocilizumab in patients with rheumatoid arthritis: results from the Swedish biologics register," *Rheumatology (Oxford),* vol. 54, pp. 1186-93, Jul 2015.

[88] R. J. Carroll, W. K. Thompson, A. E. Eyler, A. M. Mandelin, T. Cai, R. M. Zink, *et al.*, "Portability of an algorithm to identify rheumatoid arthritis in electronic health records," *Journal of the American Medical Informatics Association,* vol. 19, pp. e162-e169, 2012.

[89] B. Bruce and J. F. Fries, "The Stanford health assessment questionnaire: dimensions and practical applications," *Health and quality of life outcomes,* vol. 1, p. 20, 2003.

[90] T. Heiberg, T. K. Kvien, Ø. Dale, P. Mowinckel, G. J. Aanerud, A. B. Songe-Møller , *et al.*, "Daily health status registration (patient diary) in patients with rheumatoid arthritis: A comparison between personal digital assistant and paperpencil format," *Arthritis Care & Research,* vol. 57, pp. 454-460, 2007.

[91] C. P. Friedman, A. K. Wong, and D. Blumenthal, "Achieving a nationwide learning health system," *Science translational medicine,* vol. 2, pp. 57cm29-57cm29, 2010.

[92] K. Forslind, M. Ahlmén, K. Eberhardt, I. Hafström, and B. Svensson, "Prediction of radiological outcome in early rheumatoid arthritis in clinical practice: role of antibodies to citrullinated peptides (anti-CCP)," *Annals of the rheumatic diseases,* vol. 63, pp. 1090-1095, 2004.

[93] J. A. Wessels, S. M. van der Kooij, S. le Cessie, W. Kievit, P. Barerra, C. F. Allaart, *et al.*, "A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis," *Arthritis & Rheumatism,* vol. 56, pp. 1765-1775, 2007.

[94] L. J. Ledwich, T. M. Harrington, W. T. Ayoub, J. A. Sartorius, and E. D. Newman, "Improved influenza and pneumococcal vaccination in rheumatology patients taking immunosuppressants using an electronic health record best practice alert," *Arthritis Care & Research: Official Journal of the American College of Rheumatology,* vol. 61, pp. 1505-1510, 2009.

[95] B. Combe, I. Logeart, M. C. Belkacemi, S. Dadoun, T. Schaeverbeke, J. P. Daures, *et al.*, "Comparison of the long-term outcome for patients with rheumatoid arthritis with persistent moderate disease activity or disease remission during the first year after diagnosis: data from the ESPOIR cohort," *Ann Rheum Dis,* vol. 74, pp. 724-9, Apr 2015.

[96] E. Nikiphorou, S. Norton, A. Young, L. Carpenter, J. Dixey, D. A. Walsh, *et al.*, "Association between rheumatoid arthritis disease activity, progression of functional limitation and long-term risk of orthopaedic surgery: combined analysis of two prospective cohorts supports EULAR treat to target DAS thresholds," *Ann Rheum Dis,* vol. 75, pp. 2080-2086, Dec 2016.

[97] N. Hurst, P. Kind, D. Ruta, M. Hunter, and A. Stubbings, "Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D)," *British journal of rheumatology,* vol. 36, pp. 551-559, 1997.

[98] J. Fransen, P. Welsing, R. De Keijzer, and P. Van Riel, "Disease activity scores using C-reactive protein: CRP may replace ESR in the assessment of RA disease activity," *Ann Rheum Dis,* vol. 62, p. 151, 2004.
[99] J. Smolen, F. Breedveld, M. Schiff, J. Kalden, P. Emery, G. Eberl, *et al.*, "A simplified disease activity index for rheumatoid arthritis for use in clinical practice," *Rheumatology,* vol. 42, pp. 244-257, 2003.
[100] D. Aletaha, V. P. Nell, T. Stamm, M. Uffmann, S. Pflugbeil, K. Machold, *et al.*, "Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score," *Arthritis research & therapy,* vol. 7, p. R796, 2005.
[101] K. Arnold, J. Gosling, D. Holmes, and D. Holmes, *The Java programming language* vol. 2: Addison-wesley Reading, 2000.
[102] R. Johnson, J. Hoeller, K. Donald, C. Sampaleanu, R. Harrop, T. Risberg, *et al.*, "The spring framework-reference documentation," *interface,* vol. 21, p. 27, 2004.
[103] B. Lawson and R. Sharp, *Introducing html5*: New Riders, 2011.
[104] E. Pelegrí-Llopart and L. Cable, "JavaServer Pages™ Specification," *Sun Microsystems,* 1999.
[105] S. Bayern, *JSTL in Action*: Manning Publications Co., 2002.
[106] M. Shah, "Responsive web development using the Twitter Bootstrap framework," 2015.
[107] M. Mathworks, "MATLAB and Statistics Toolbox Release. Natick, MA: The MathWorks," ed: Inc, 2017.
[108] I. Bratko, *Prolog programming for artificial intelligence*: Pearson education, 2001.
[109] K. Sagonas, T. Swift, and D. S. Warren, "XSB as an efficient deductive database engine," *ACM SIGMOD Record,* vol. 23, pp. 442-453, 1994.
[110] G. Antoniou and A. Bikakis, "Dr-prolog: A system for defeasible reasoning with rules and ontologies on the semantic web," *IEEE transactions on knowledge and data engineering,* vol. 19, pp. 233-245, 2006.
[111] M. L. Prevoo, M. A. van 't Hof, H. H. Kuper, M. A. van Leeuwen, L. B. van de Putte, and P. L. van Riel, "Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis," *Arthritis Rheum,* vol. 38, pp. 44-8, Jan 1995.
[112] O. Brocq, E. Millasseau, C. Albert, C. Grisot, P. Flory, C. H. Roux, *et al.*, "Effect of discontinuing TNFalpha antagonist therapy in patients with remission of rheumatoid arthritis," *Joint Bone Spine,* vol. 76, pp. 350-5, Jul 2009.
[113] D. E. Furst, A. L. Pangan, L. R. Harrold, H. Chang, G. Reed, J. M. Kremer, *et al.*, "Greater likelihood of remission in rheumatoid arthritis patients treated earlier in the disease course: results from the Consortium of Rheumatology Researchers of North America registry," *Arthritis Care Res (Hoboken),* vol. 63, pp. 856-64, Jun 2011.
[114] J. T. Einarsson, P. Geborek, T. Saxne, and M. C. Kapetanovic, "Sustained Remission in Tumor Necrosis Factor Inhibitor–treated Patients with Rheumatoid Arthritis: A Population-based Cohort Study," *The Journal of rheumatology,* vol. 42, pp. 741-748, 2015.
[115] E. Balogh, J. Madruga Dias, C. Orr, R. Mullan, L. Harty, O. FitzGerald, *et al.*, "Comparison of remission criteria in a tumour necrosis factor inhibitor treated rheumatoid arthritis longitudinal cohort: patient global health is a confounder," *Arthritis Res Ther,* vol. 15, p. R221, Dec 24 2013.
[116] C. Barnabe, J. Homik, S. G. Barr, L. Martin, and W. P. Maksymowych, "The effect of different remission definitions on identification of predictors of both point and

sustained remission in rheumatoid arthritis treated with anti-TNF therapy," *J Rheumatol,* vol. 41, pp. 1607-13, Aug 2014.
[117] Y. Tanaka, S. Hirata, S. Kubo, S. Fukuyo, K. Hanami, N. Sawamukai, *et al.*, "Discontinuation of adalimumab after achieving remission in patients with established rheumatoid arthritis: 1-year outcome of the HONOR study," *Ann Rheum Dis,* vol. 74, pp. 389-95, Feb 2015.
[118] P. Hamann, R. Holland, K. Hyrich, J. D. Pauling, G. Shaddick, A. Nightingale, *et al.*, "Factors Associated With Sustained Remission in Rheumatoid Arthritis in Patients Treated With Anti–Tumor Necrosis Factor," *Arthritis care & research,* vol. 69, pp. 783-793, 2017.
[119] I. D. Flouri, T. E. Markatseli, K. A. Boki, I. Papadopoulos, F. N. Skopouli, P. V. Voulgari, *et al.*, "Comparative Analysis and Predictors of 10-year Tumor Necrosis Factor Inhibitors Drug Survival in Patients with Spondyloarthritis: First-year Response Predicts Longterm Drug Persistence," *J Rheumatol,* vol. 45, pp. 785-794, Jun 2018.
[120] R. Landewé, D. van der Heijde, L. Klareskog, R. van Vollenhoven, and S. Fatenejad, "Disconnect between inflammation and joint destruction after treatment with etanercept plus methotrexate: results from the trial of etanercept and methotrexate with radiographic and patient outcomes," *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology,* vol. 54, pp. 3119-3125, 2006.
[121] J. Smolen, C. Han, M. Bala, R. Maini, J. Kalden, D. Van der Heijde, *et al.*, "Evidence of radiographic benefit of treatment with infliximab plus methotrexate in rheumatoid arthritis patients who had no clinical improvement: a detailed subanalysis of data from the Anti–Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy study," *Arthritis & Rheumatism,* vol. 52, pp. 1020-1030, 2005.
[122] J. S. Smolen, J. C. M. Avila, and D. Aletaha, "Tocilizumab inhibits progression of joint damage in rheumatoid arthritis irrespective of its anti-inflammatory effects: disassociation of the link between inflammation and destruction," *Annals of the rheumatic diseases,* vol. 71, pp. 687-693, 2012.
[123] B. Mehta, S. Pedro, G. Ozen, A. Kalil, F. Wolfe, T. Mikuls, *et al.*, "Serious infection risk in rheumatoid arthritis compared with non-inflammatory rheumatic and musculoskeletal diseases: a US national cohort study," *RMD open,* vol. 5, p. e000935, 2019.
[124] K. Au, G. Reed, J. R. Curtis, J. M. Kremer, J. D. Greenberg, V. Strand, *et al.*, "High disease activity is associated with an increased risk of infection in patients with rheumatoid arthritis," *Annals of the rheumatic diseases,* vol. 70, pp. 785-791, 2011.
[125] A. Weaver, O. Troum, M. Hooper, A. S. Koenig, S. Chaudhari, J. Feng, *et al.*, "Rheumatoid arthritis disease activity and disability affect the risk of serious infection events in RADIUS 1," *The Journal of Rheumatology,* vol. 40, pp. 1275-1281, 2013.
[126] N. A. Accortt, T. Lesperance, M. Liu, S. Rebello, M. Trivedi, Y. Li, *et al.*, "Impact of sustained remission on the risk of serious infection in patients with rheumatoid arthritis," *Arthritis care & research,* vol. 70, pp. 679-684, 2018.
[127] E. E. Arts, J. Fransen, A. A. den Broeder, C. D. Popa, and P. L. van Riel, "The effect of disease duration and disease activity on the risk of cardiovascular disease in rheumatoid arthritis patients," *Annals of the rheumatic diseases,* vol. 74, pp. 998-1003, 2015.
[128] E. E. Arts, J. Fransen, A. A. Den Broeder, P. L. van Riel, and C. D. Popa, "Low disease activity (DAS28≤ 3.2) reduces the risk of first cardiovascular event in rheumatoid arthritis: a time-dependent Cox regression analysis in a large cohort study," *Annals of the Rheumatic Diseases,* vol. 76, pp. 1693-1699, 2017.

[129] H. Tsuji, K. Yano, M. Furu, N. Yamakawa, K. Ikari, M. Hashimoto, *et al.*, "Time-averaged disease activity fits better joint destruction in rheumatoid arthritis," *Sci Rep*, vol. 7, p. 5856, Jul 19 2017.
[130] P. Kiely, D. Walsh, R. Williams, and A. Young, "Outcome in rheumatoid arthritis patients with continued conventional therapy for moderate disease activity—the early RA network (ERAN)," *Rheumatology*, vol. 50, pp. 926-931, 2010.

B. Appendix – Supplementary Tables

Supplementary Table 1. Sequential bDMARDs Treatments (BTs) of Cohort and Patient Groups LDA, MDA, and HDA.

BTs	Cohort			LDA			MDA			HDA		
	1st BT	2nd BT	3rd BT	1st BT	2nd BT	3rd BT	1st BT	2nd BT	3rd BT	1st BT	2nd BT	3rd BT
Samples	527	334	203	90	28	7	295	180	92	142	126	104
Adalimu-mab	124 (24)	70 (21)	15 (7)	27 (30)	9* (32)	0 (0)	69 (23)	36 (20)	6 (7)	28 (18)	25 (20)	9 (9)
Etane-rcept	70 (13)	71* (21)	24 (12)	8 (9)	4 (14)	2 (29)	45 (15)	41* (23)	11 (12)	17 (12)	26* (21)	11 (11)
Inflixi-mab	264* (50)	39 (12)	7 (3)	53* (59)	2 (7)	0 (0)	143* (49)	23 (13)	2 (2)	68* (48)	14 (11)	5 (5)
Certoli-zumab	4 (0.7)	3 (0.9)	1 (0.5)	0 (0)	1 (4)	0 (0)	3 (1)	2 (1)	0 (0)	1 (1)	0 (0)	1 (1)
Golimu-mab	5 (0.9)	5 (2)	6 (3)	0 (0)	1 (4)	0 (0)	2 (1)	3 (2)	4 (4)	3 (2)	1 (1)	2 (2)
Tocilizu-mab	10 (0.2)	30 (9)	42 (21)	0 (0)	3 (11)	0 (0)	8 (3)	17 (9)	21 (23)	2 (2)	10 (8)	21 (20)

Anakinra	9 (0.2)	6 (2)	6 (3)	0 (0)	0 (0)	0 (0)	2 (1)	3 (2)	2 (2)	7 (5)	3 (2)	4 (4)
Rituximab	13 (2)	55 (16)	55* (27)	1 (1)	7 (25)	4* (57)	8 (3)	25 (14)	23* (25)	4 (3)	23 (18)	28* (27)
Abatacept	28 (5)	55 (16)	47 (23)	1 (1)	1 (4)	1 (14)	15 (5)	30 (17)	23* (25)	12 (9)	24 (19)	23 (22)

§ Results are presented as counts n (%) of biologic drug agents. LDA= Persistent Low Disease; MDA= Persistent Medium Disease Activity; HDA = Persistent High Disease.

* The most frequent biologic drug agent as first, second and third sequential treatment, in individual patients group.

C. Appendix – Defeasible Logic Metaprogram

On the TuProlog reasoning engine it is loaded the DR-PROLOG metaprogram. This metaprogram is actually a prolog program, implementing the defeasible logic, in a shorter more lightweight version of the original the ideas which are described in [48].

```
supportive_rule(Name, Head, Body) :- strict(Name, Head, Body).
supportive_rule(Name, Head, Body) :- defeasible(Name, Head, Body).
rule(Name,Head,Body) :- supportive_rule(Name, Head, Body).

definitely(X):- fact(X).
definitely(X):- strict(R,X,L), definitely_provable(L).
definitely(X):- strict0(R,X,L), definitely_provable(L).
definitely(neg(X)):- strict1(R,neg(X),L), definitely_provable(L), not(definitely(X)).
definitely(X):- strict2(R,X,L), definitely_provable(L).

definitely_provable([]).
definitely_provable(X):- definitely(X).
definitely_provable([X1|X2]):- definitely_provable(X1), definitely_provable(X2).

defeasibly(X):- definitely(X).
```

```
defeasibly(X):- negation(X,X1), supportive_rule(R,X,L), defeasibly_provable(L), not(definitely(X1)),
not(overruled(R,X)).
defeasibly_provable([]).
defeasibly_provable(X):- defeasibly(X).
defeasibly_provable([X1|X2]):- defeasibly_provable(X1), defeasibly_provable(X2).

overruled(R,X):- negation(X,X1), supportive_rule(S,X1,U), defeasibly_provable(U),
not(defeated(S,X1)).
defeated(S,X):- superior(T,S), negation(X,X1), supportive_rule(T,X1,V), defeasibly_provable(V).

negation(~(X),X):- !.
negation(X,~(X)).

append([],List,List).
append([Head|Tail],List2,[Head|Result]):- append(Tail,List2,Result).
member(N,[N|Tail]).
member(N,[I|Tail]):- member(N,Tail).

minus_set([E|X],Y,Z):- member(E,Y),minus_set(X,Y,Z),!.
minus_set([E|X],Y,[E|Z]):-minus_set(X,Y,Z),not(member(E,Y)), !.
minus_set([],Y,[]).
strict(e,w,r).
defeasible(y,t,e).
fact(w).

superior(e,w).
```